VOLUME 103 PART B NUMBER 9 MAY 1956

The Proceedings

OF

THE INSTITUTION OF ELECTRICAL ENGINEERS

FOUNDED 1871: INCORPORATED BY ROYAL CHARTER 1921

PART B

RADIO AND ELECTRONIC ENGINEERING
(INCLUDING COMMUNICATION ENGINEERING)

SAVOY PLACE . LONDON W.C.2

Price Seven Shillings and Sixpence

The Institution of Electrical Engineers

FOUNDED 1871
INCORPORATED BY ROYAL CHARTER 1921

PATRON: HER MAJESTY THE QUEEN

COUNCIL 1955-56

President

SIR GEORGE H. NELSON, Bart.

Past-Presidents

SIR JAMES SWINBURNE, Bart., F.R.S.
W. H. ECCLES, D.Sc., F.R.S.
THE RT. HON. THE EARL OF MOUNT EDGCUMBE, T.D.
J. M. DONALDSON, M.C.
PROFESSOR E. W. MARCHANT, D.Sc.
P. V. HUNTER, C.B.E.
H. T. YOUNG.
SIR GEORGE LEE, O.B.E., M.C.
SIR ARTHUR P. M. FLEMING, C.B.E., D.Eng., LL.D.
J. R. BEARD, C.B.E., M.Sc.
SIR NOEL ASHBRIDGE, B.Sc.(Eng.).
COLONEL SIR A. STANLEY ANGWIN, K.B.E., D.S.O., M.C., T.D., D.Sc. (Eng.).
SIR HARRY RAILING, D.Eng.
P. DUNSHEATH, C.B.E., M.A., D.Sc. (Eng.).
SIR VINCENT Z. DE FERRANTI, M.C.
T. G. N. HALDANE, M.A.
PROFESSOR E. B. MOULLIN, M.A., Sc.D.
SIR ARCHIBALD J. GILL, B.Sc.(Eng.).
SIR JOHN HACKING.
COLONEL B. H. LEESON, C.B.E., T.D.
SIR HAROLD BISHOP, C.B.E., B.Sc.(Eng.).
J. ECCLES, C.B.E., B.Sc.

Vice-Presidents

T. E. GOLDUP, C.B.E.
S. E. GOODALL, M.Sc.(Eng.).
SIR HAMISH D. MACLAREN, K.B.E., C.B., D.F.C., LL.D., B.Sc.
WILLIS JACKSON, D.Sc., D.Phil., Dr.Sc.Tech., F.R.S.
SIR W. GORDON RADLEY, C.B.E., Ph.D.(Eng.).

Honorary Treasurer

THE RT. HON. THE VISCOUNT FALMOUTH

Ordinary Members of Council

PROFESSOR H. E. M. BARLOW, Ph.D., B.Sc.(Eng.).
J. BENNETT.
C. M. COCK.
A. R. COOPER, M.Eng.
A. T. CRAWFORD, B.Sc.
B. DONKIN, B.A.
PROFESSOR J. GREIG, M.Sc., Ph.D.
F. J. LANE, O.B.E., M.Sc.
G. S. C. LUCAS, O.B.E.
D. MCDONALD, B.Sc.
C. T. MELLING, C.B.E., M.Sc.Tech.
H. H. MULLENS, B.Sc.
W. F. PARKER.
R. L. SMITH-ROSE, C.B.E., D.Sc., Ph.D.
G. L. WATES, J.P.
G. O. WATSON.
D. B. WELBOURN, M.A.
J. H. WESTCOTT, B.Sc.(Eng.), Ph.D.
E. L. E. WHEATCROFT, M.A.
R. T. B. WYNN, C.B.E., M.A.

Chairman and Past-Chairmen of Sections

Measurement and Control:
W. BAMFORD, B.Sc.
*M. WHITEHEAD.

Radio and Telecommunication:
H. STANESBY.
*C. W. OATLEY, O.B.E., M.A., M.Sc.

Supply:
L. DRUCQUER.
*J. D. PEATTIE, B.Sc.

Utilization:
D. B. HOGG, M.B.E., T.D.
*J. I. BERNARD, B.Sc.Tech.

Chairmen and Past-Chairmen of Local Centres

East Midland Centre:
F. R. C. ROBERTS.
*J. M. MITCHELL, B.Sc., Ph.D.

Mersey and North Wales Centre:
PROFESSOR J. M. MEEK, D.Eng.
*P. R. DUNN, B.Sc.

North Midland Centre:
F. BARRELL.
*G. CATON.

North-Eastern Centre:
A. H. KENYON.
*G. W. B. MITCHELL, B.A.

Southern Centre:
L. H. FULLER, B.Sc.(Eng.).
*E. A. LOGAN, M.Sc.

North-Western Centre:
G. V. SADLER.
*PROFESSOR E. BRADSHAW, M.B.E., M.Sc.Tech., Ph.D.

Northern Ireland Centre:
MAJOR E. N. CUNLIFFE, B.Sc.Tech.
*MAJOR P. L. BARKER, B.Sc.

Western Centre:
T. G. DASH, J.P.
*A. N. IRENS.

Scottish Centre:
*E. WILKINSON, Ph.D., B.Eng.
*J. S. HASTIE, B.Sc.(Eng.).

South Midland Centre:
H. S. DAVIDSON, T.D.
*A. R. BLANDFORD.

* Past-Chairman.

RADIO AND TELECOMMUNICATION SECTION COMMITTEE 1955-56

Chairman

H. STANESBY

Vice-Chairmen

R. C. G. WILLIAMS, Ph.D., B.Sc.(Eng.).
J. S. MCPETRIE, Ph.D., D.Sc.

Past-Chairmen

C. W. OATLEY, O.B.E., M.A., M.Sc.
J. A. SMALE, C.B.E., A.F.C., B.Sc.

Ordinary Members of Committee

PROF. H. E. M. BARLOW, Ph.D., B.Sc.(Eng.).
F. S. BARTON, C.B.E., M.A., B.Sc.
A. J. BIGGS, Ph.D., B.Sc.
E. V. D. GLAZIER, Ph.D.(Eng.), B.Sc.
G. G. MACFARLANE, Dr.Ing., B.Sc.
B. N. MACLARTY, O.B.E.
H. PAGE, M.Sc.
W. ROSS, M.A.
L. RUSHFORTH, M.B.E., B.Sc.
T. B. D. TERRONI, B.Sc.
A. M. THORNTON, B.Sc.
F. WILLIAMS, B.Sc.

And

The President (*ex officio*).
The Chairman of the Papers Committee.
PROF. H. E. M. BARLOW, Ph.D., B.Sc.(Eng.) (representing the Council).
E. H. COOKE-YARBOROUGH (Co-opted Member).
BRIG. E. J. H. MOPPETT (representing the Cambridge Radio and Telecommunication Group).
J. MOIR (representing the South Midland Radio and Telecommunication Group).
D. H. THOMAS, M.Sc.Tech., B.Sc.(Eng.) (representing the North-Eastern Radio and Measurements Group).

The following nominees of Government Departments:
Admiralty: CAPTAIN G. C. F. WHITAKER, R.N.
Air Ministry: AIR COMMODORE G. H. RANDLE, R.A.F., B.A.
Department of Scientific and Industrial Research: B. G. PRESSEY, M.Sc.(Eng.), Ph.D.
Ministry of Supply: BRIG. J. D. HAIGH, O.B.E., M.A.
Post Office: CAPTAIN C. F. BOOTH, O.B.E.
War Office: COL. E. I. E. MOZLEY, M.A.

Secretary

W. K. BRASHER, C.B.E., M.A., M.I.E.E.

Deputy Secretary

F. JERVIS SMITH M.I.E.E.

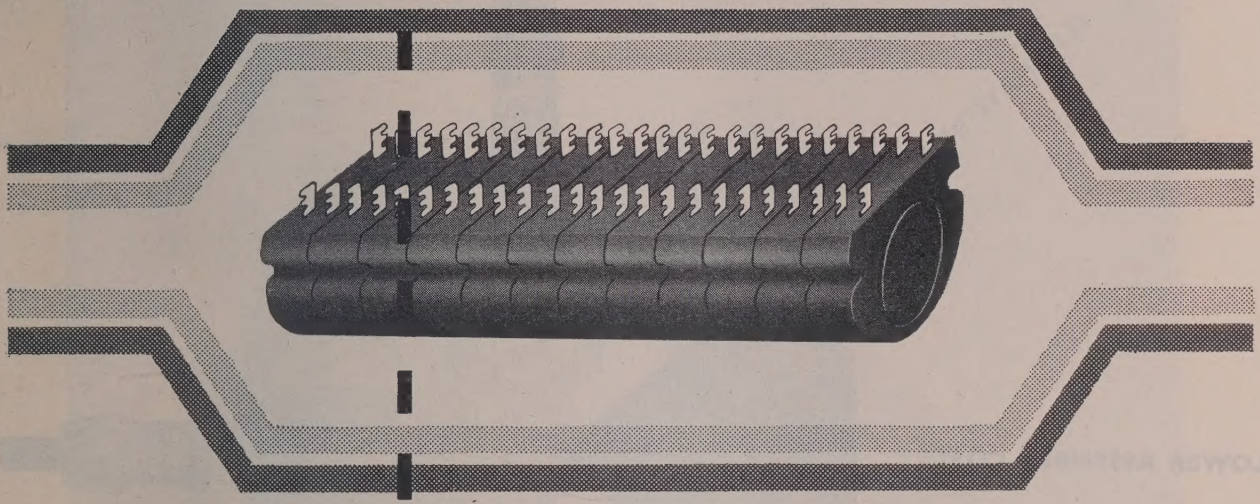
Assistant Secretary

F. C. HARRIS.

Editor-in-Chief

G. E. WILLIAMS, B.Sc.(Eng.), M.I.E.E.

Loading coils **inside** the cable splice

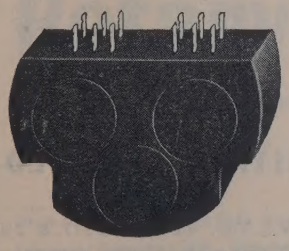


The advantages of the splice loading technique are particularly marked in the loading of small cables of up to 74 pairs. The coils can be included in a jointing sleeve or unit of only slightly larger diameter than would normally be used.

The loading coils in the Mullard L.160 Series are designed specifically for this technique. They are cast in resin, which provides complete protection from climatic conditions and allows a telephone administration to store them ready for building into loading units as and when required. Both single and triple assemblies are available for different sizes of cable.

Ferroxcube pot cores give these coils certain electrical advantages over conventional types, particularly in the loading of higher frequency circuits such as those encountered in programme and carrier applications.

You are invited to write for leaflets describing the Mullard L.160 Series coils and simple units for pole and splice loading.



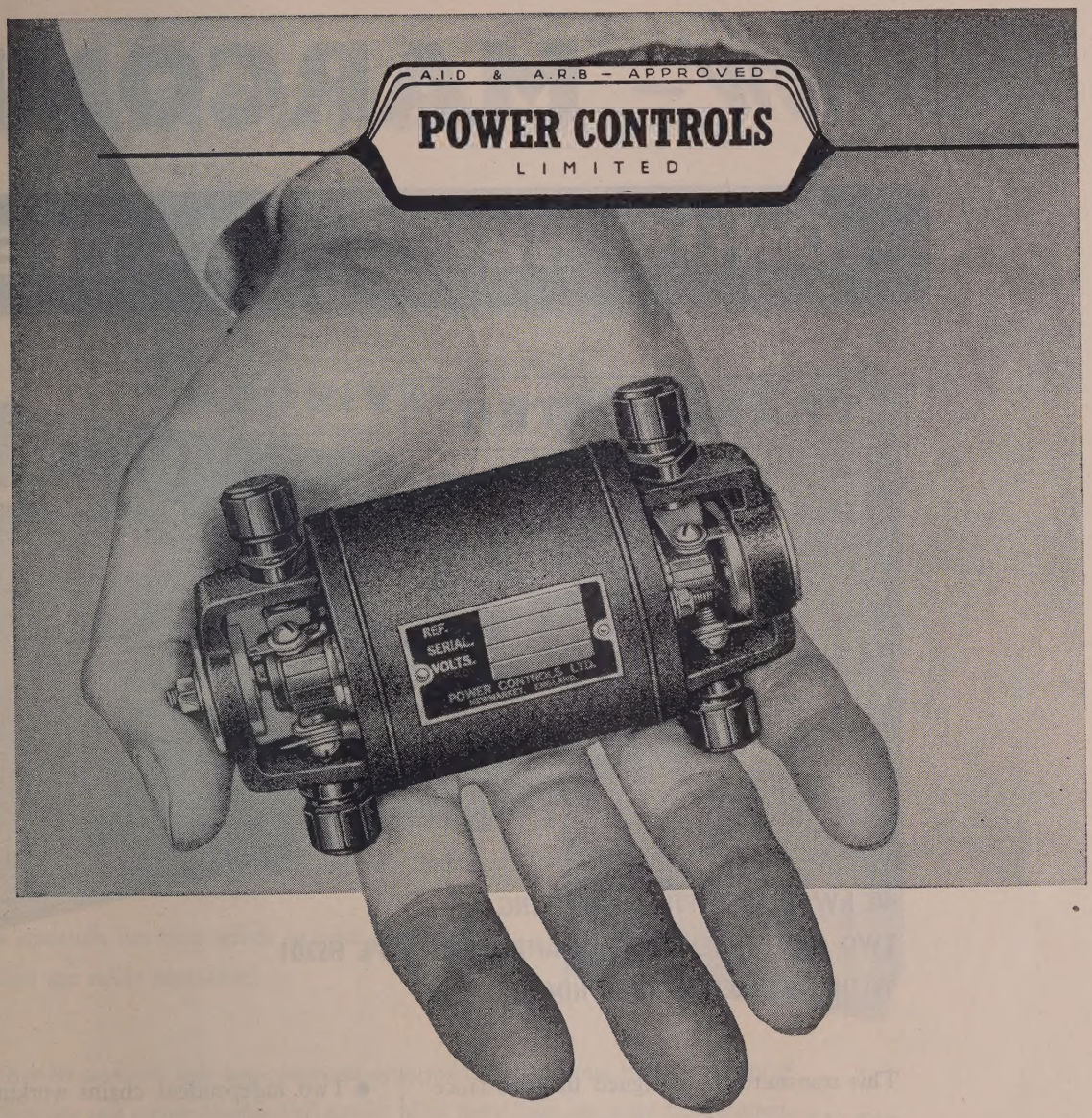
TRIPLE COIL ASSEMBLY

Mullard

SPECIALISED ELECTRONIC EQUIPMENT

MULLARD LTD · EQUIPMENT DIVISION
CENTURY HOUSE · SHAFTESBURY AVENUE · W.C.2

(MI453A)



Rotary Transformers

Have you a transformer problem? If so, we can help you. We can undertake to develop and manufacture rotary transformers to your specification.

The illustration shows a typical transformer which we are manufacturing for a specific requirement. Made for 6, 12 or 24 volts D.C. input, it can supply a continuous D.C. output of 350 volts at 30 mA. or an intermittent output of 310 volts at 60 mA. The no-load current consumption is 2.2 amps. at 11.5 volts and the ripple voltage is less than 6 volts r.m.s. on 60 mA. load. The size is only 4-9/16″ long by 2-21/32″ across the brush terminals.

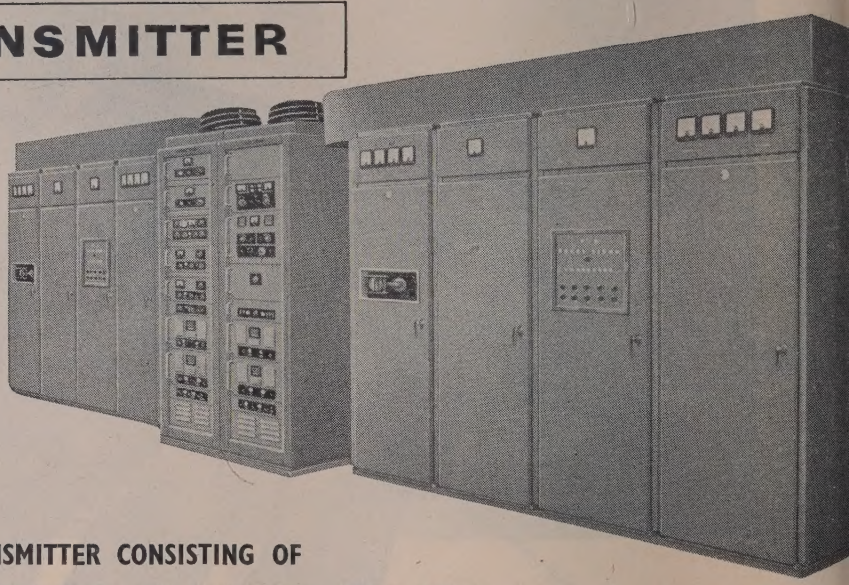
Power Controls Ltd., Exning Road, Newmarket, Suffolk

Telephone: Newmarket 3181. Telegrams: Powercon, Newmarket

Now— MARCONI'S

IONOSPHERIC SCATTER

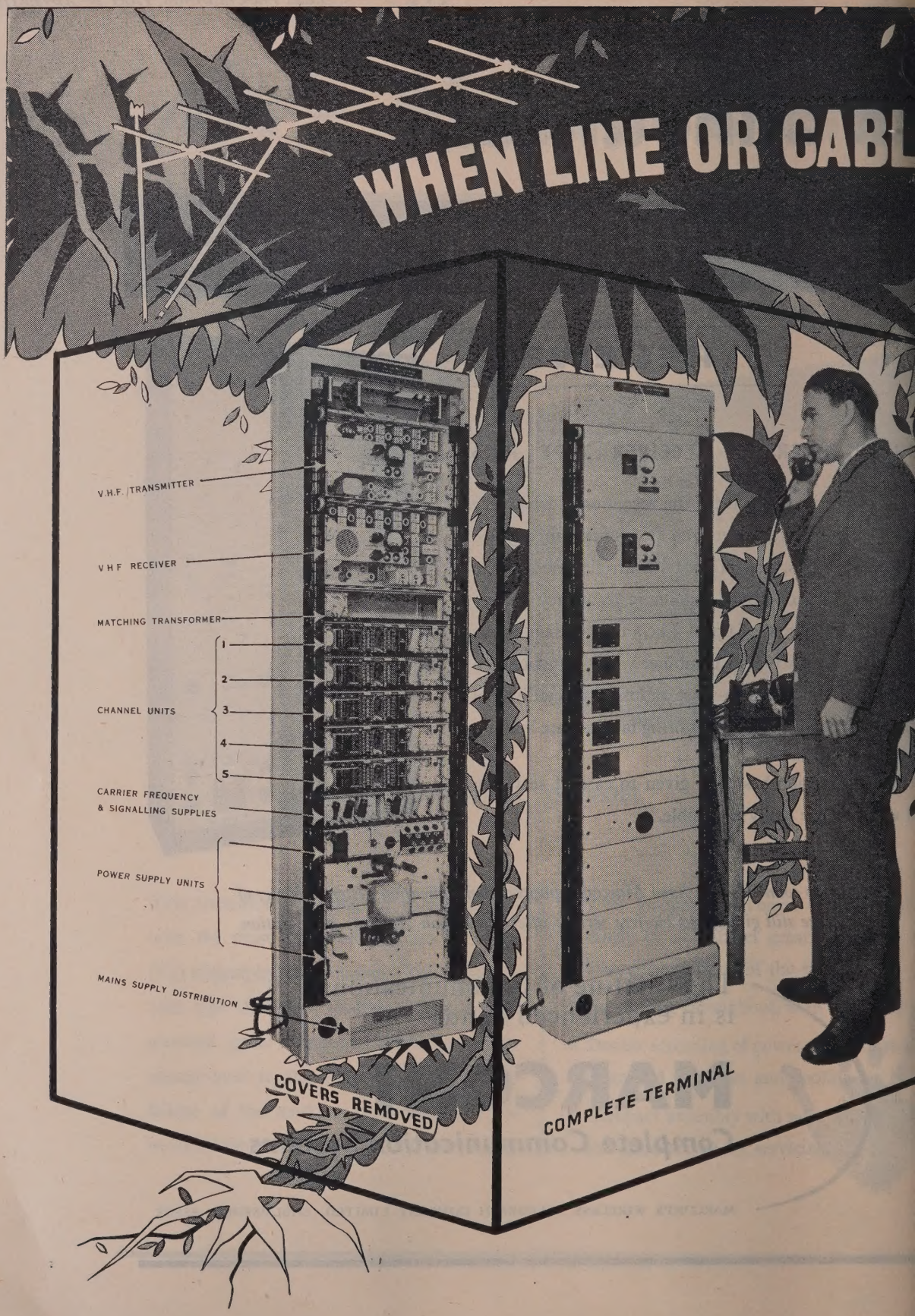
TRANSMITTER



40 kW TRANSMITTER CONSISTING OF TWO INDEPENDENT 20 kW AMPLIFIERS TYPE HS201 WITH FSK EXCITER TYPE HD65

This transmitter is designed in accordance with the most advanced practice for FSK (F1) telegraphy transmission. The FSK exciter type HD65 is a separate unit; it has the essential parts duplicated, with automatic change-over to the standby equipment on failure of the working part. Facilities are included for monitoring the FSK waveforms.

- Two independent chains working indi dually or in parallel greatly improve t inherent reliability of the system.
- Air cooling throughout, with dust filteri
- Double screening of power stages to redu unwanted radiation and cooling-air noi
- Compact assembly with safety interlocki and good access for servicing.



THE GENERAL ELECTRIC COMPANY LIMITED OF ENGLA

5 - CIRCUIT

UNCTION RADIO EQUIPMENT

L—INEXPENSIVE—AVAILABLE FOR QUICK DELIVERY—this equipment provides up to five uits by frequency-division-multiplex operation over the frequency-modulated VHF o link and has the advantage of narrow occupancy in the congested VHF band.

RATING over a distance of 50 miles the total cost of a system consisting of two complete inal equipments, aerials and feeder cables is less than the cost of the copper wire required erive a single open-wire pair for the provision of equivalent facilities.

panels of the equipment are of the slide-in type and can be removed in a few seconds for of maintenance.

For further information write for Standard Specification SPO 5050

OTHER G.E.C. TRANSMISSION EQUIPMENT INCLUDES:

e circuit, 3- and 12-circuit equipment for operation pen-wire lines.
60-circuit equipment for operation over non-loaded r cables.
ial line equipment for television and telephony.
Power line carrier equipment.
Coaxial cable terminal equipment for television (405, 525 and 625-line systems) and telephony (960 circuits).
Wide band UHF trunk radio systems for television and telephony up to 720 circuits.

EVERYTHING FOR TELECOMMUNICATIONS BY OPEN-WIRE LINE, CABLE AND RADIO, SINGLE OR MULTI-CIRCUIT, OR T.V. LINK. SHORT, MEDIUM, OR LONG HAUL.

LEPHONE, RADIO AND TELEVISION WORKS, COVENTRY

2 new Stabilised Power Supply Units

Three stabilised voltage outputs are provided by each of the two new Ediswan Power Supply Units Types R.1260 and R.1280. In each case two major D.C. outputs variable from 0 to 300 v. are provided, plus a low current–D.C. output of reverse polarity.

These units are very suitable power sources for experimental electronic circuits and can frequently be used in place of several single power supplies.

Both units work from normal A.C. mains supply and are designed for standard 19 in. rack mounting or for bench use. Meters are included.

SPECIFICATION

Input

200–250 v., 40–100 C.P.S.

Outputs

(1) High stability D.C. output 0 to 300 v. in one range.
Maximum current: type R.1260 75mA.
type R.1280 50mA.

(2) High stability D.C. output 0 to 300 v. in one range.
Maximum current (both types) 75 mA.

(3) High stability D.C. output-200 v. non-variable.
Maximum current (both types) 25 mA.

(4) and (5) Two independent 6·3 v. 4 amp. A.C. unstabilised ouputs, both centre tapped.

Notes:

(A) The negative terminals of the two 300 v. supplies and the positive terminal of the 200 v. supply are common;

(B) Either output 1 or 2 may be taken with output 3 to give a supply of 200–500 v. at 25 mA.

Stability

With change of mains input voltage (outputs 1, 2 and 3);

A change of 10 v. in the mains supply results in an output change of less than 0·1 v.

With change of load: A change from zero to full load results in output voltage changes as follows:—

Outputs 1 and 2	Less than 0·3 v.
Output 3	Less than 0·1 v.

Output Resistance (Outputs 1, 2 and 3).
Less than 3 ohms.

Ripple

Outputs 1 and 2	Less than 2 mV. R.M.S.
Output 3	Less than 4 mV. R.M.S.

Output Circuits

All outputs isolated from earth, and any point may be connected to earth. All terminals can be worked at up to 500 volts from earth. Outputs 1 and 2 may be switched without affecting any of the heater supplies.

Meters

Each 300 v. supply has a meter which can be switched to be a voltmeter or milli-ammeter.

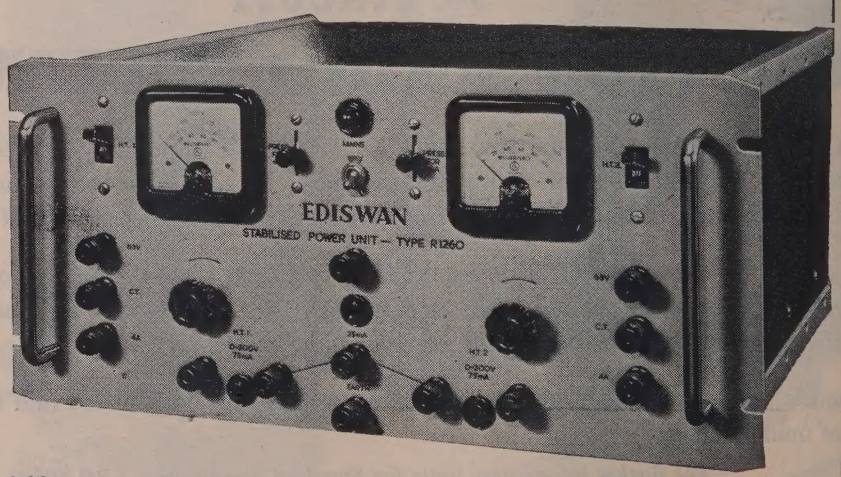
Mounting

R.1260 and R.1280 are both suitable for standard 19 in. rack mounting, or for bench use.

Price

R.1260	£78 nett
R.1280	£86 nett

Bench stand for R.1280, £2 nett



Further information on this and other Ediswan Stabilised Power Supply Units available on request

EDISWAN

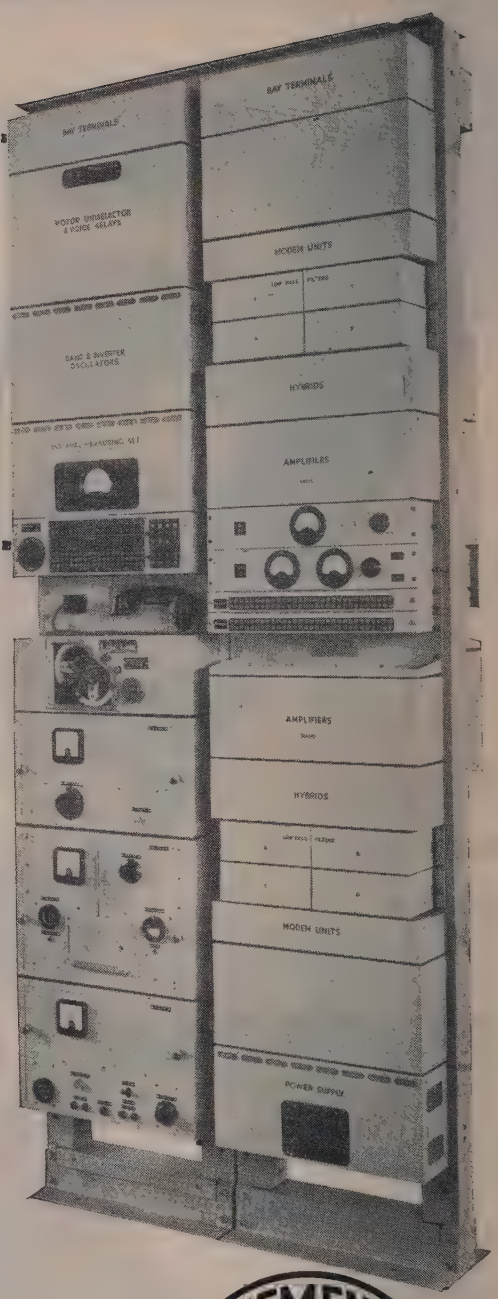
RADIO DIVISION

THE EDISON SWAN ELECTRIC CO. LTD. • 155 Charing Cross Road • London, W.C.2.
Telephone: Gerrard 8660. Member of the A.E.I. Group of Companies. Telegrams: Ediswan, Westcent, London

MARCONI–SIEMENS

Five Band Split Privacy Radio Telephone Equipment

(Type HW 12)



This equipment, which may be switched in or out of use at the radio terminal, provides a very high degree of privacy for speech on a radio-telephone circuit by:-
(1) splitting the speech band of 250-3000 c/s into five sub-bands of 550 c/s and recombining them in different relative positions,
(2) inverting the frequency range of any one or more of the sub-bands, and
(3) rearranging the combination of the sub-bands simultaneously at both ends of the radio-circuit in accordance with a pre-arranged sequence at controlled intervals between 4 and 20 seconds.
The resulting speech band, which modulates the transmitter, is unintelligible and the frequent regrouping of the sub-bands, with or without inversion precludes any simple method of interception. A reversal of the process at the distant terminal restores the original speech. The processes involved are reversible, thus common channel equipment can be used for both transmission and reception. Amplifiers in the privacy path compensate for the losses in band splitting and recombining. The simultaneous switching system, operates by means of relays under the control of a synchronous motor driven by a high precision crystal oscillator, this does away with the need for a transmitter pilot tone.

Marconi

THE LINK BETWEEN RADIO AND LINE COMMUNICATIONS

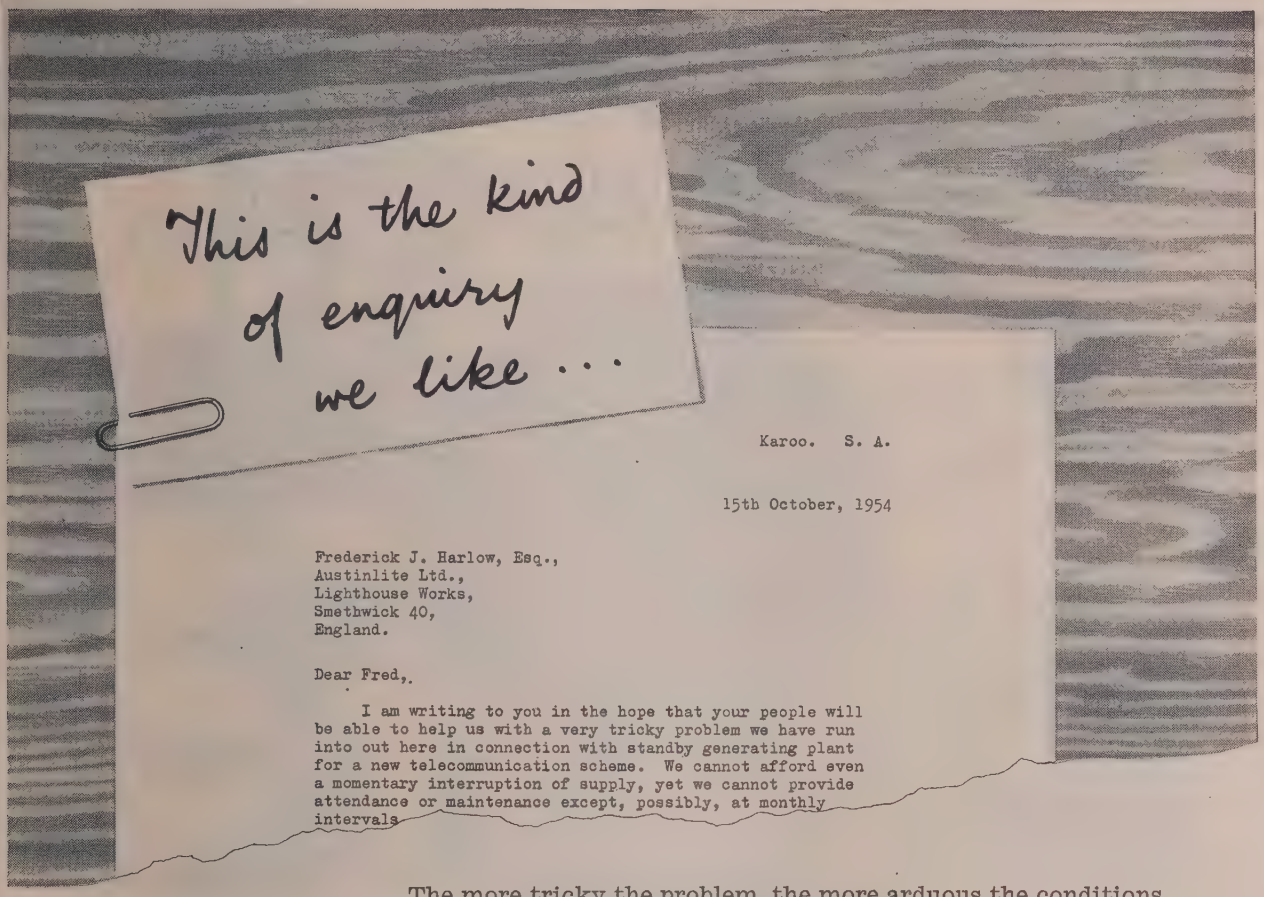
SIEMENS BROTHERS

Full technical details of this and other Marconi-Siemens equipment, which provides completely integrated radio and line telegraph and telephone systems may be obtained from either

MARCONI'S WIRELESS TELEGRAPH COMPANY LIMITED, CHELMSFORD, ESSEX
OR SIEMENS BROTHERS & CO., LIMITED, WOOLWICH, LONDON, S.E.18

MS2

Automatic Generating Plant



The more tricky the problem, the more arduous the conditions under which the equipment must operate, the better we like it. Difficult, unusual generating plant is our metier. It does not matter how long it must run without attention. It does not matter whether there is a mains supply or not. Nor if the supply is erratic and unreliable. Even if the requirements do not come within our standard range Austinlite plant can be designed which will ensure continuous, steady and extremely reliable supply.

During the past 25 years Austinlite plant has been steadily developed, always with the emphasis on quality and dependability. We can now provide automatic generating plant of many types in sizes from 1.4 to 250kVA, for continuous unattended operation, or for mains standby work — to the strictest no break specification where necessary. However difficult a particular problem may be, we have good reason for the confidence with which we affirm that Austinlite can tackle it.

In this country our plant is used by the G.P.O., the Ministry of Civil Aviation and British Railways. It is installed in many parts of the world in lighthouses where reliability is vital. It is in operation in South Africa, Australia, Denmark, the U.S.S.R., Malta, Syria, New Zealand and Norway; as well as a number of tropical countries including Nigeria, Malaya and Borneo.

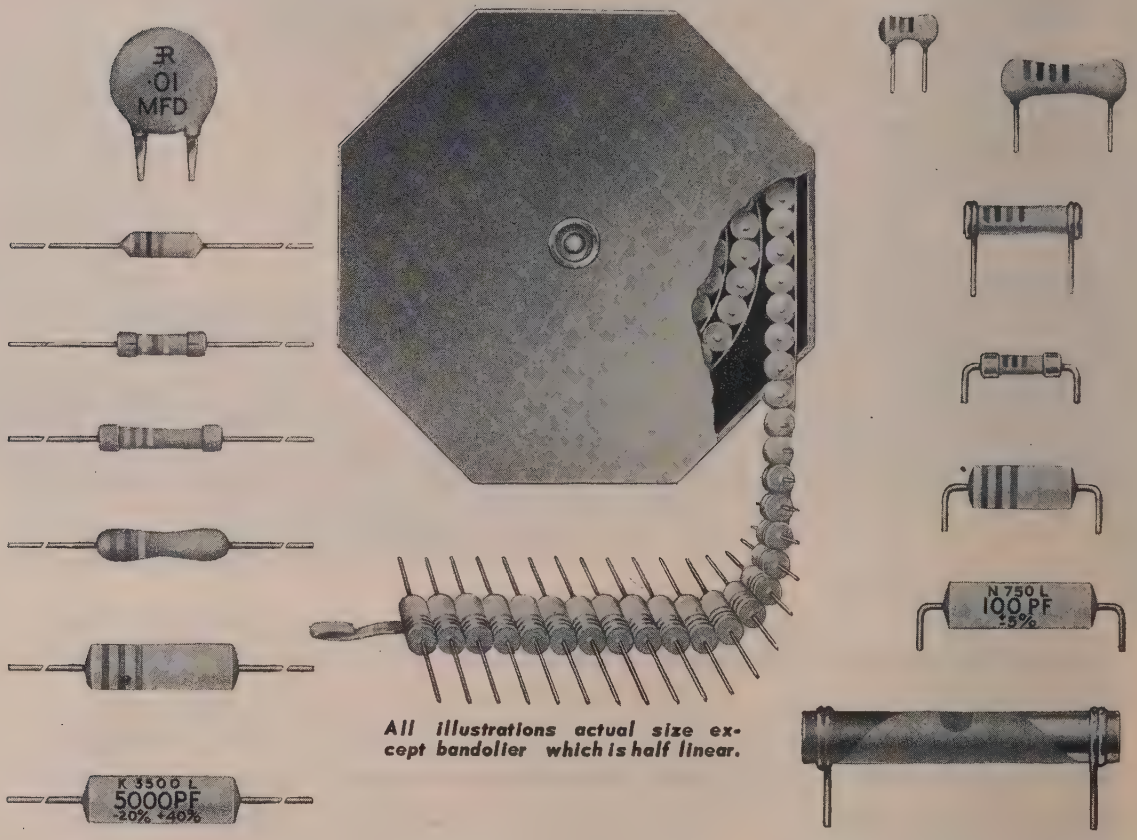
Austinlite AUTOMATIC GENERATING PLANT Tailor-made by STONE-CHANCE LTD.

(The makers of Sumo Pumps and Stone-Chance Lighthouses)

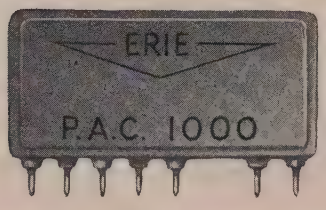
28 ST. JAMES'S SQUARE, LONDON, S.W.1.

Tel: Trafalgar 1954

the ERIE* approach to AUTOMATION...



All illustrations actual size except bandolier which is half linear.

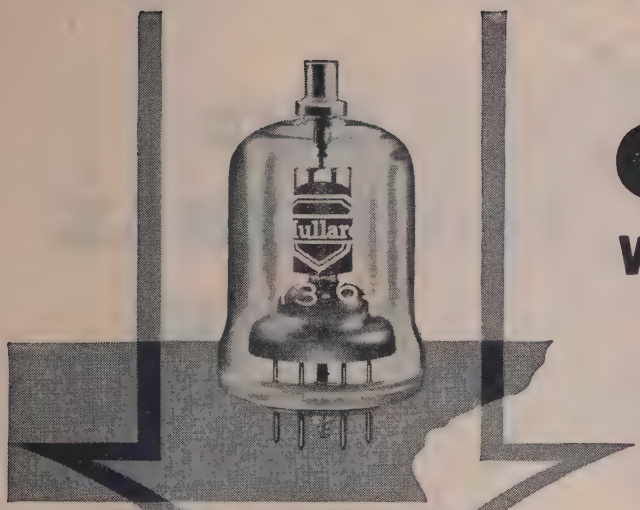


... ranges from standard resistors and capacitors, with axial leads, tape fed from bandoliers, and standard resistors and capacitors, with shortened preformed radial leads, supplied in heat sealed polythene envelopes, to packaged sub-assemblies known as the Erie 'PAC', which, in a single unit, can accommodate a prewired circuit comprising as many as ninety-two standard resistor and capacitor elements.

ERIE* *Resistor Ltd*

* Registered Trade Marks

Carlisle Road, The Hyde, London, N.W.9, England. Telephone: COLindale 8011. *Factories: London and Great Yarmouth, England; Trenton, Ont., Canada; Erie, Pa., and Holly Springs, Miss., U.S.A.*



QY3-65

V.H.F. POWER TETRODE

a new addition to a popular range

This new tetrode, the QY3-65, embodies a similar technique and construction to the Mullard range of all-glass transmitting valves which are already well established in communications and industry. It has an anode dissipation of 65 watts and a maximum frequency of 250 Mc/s and is directly interchangeable with the American 4-65A.

Relatively high outputs can be obtained from the QY3-65 at low anode voltages, and its quick heating filament allows power consumption during standby to be reduced to a minimum.
Write for detailed information on this valve, power triodes and other tetrodes made by Mullard.

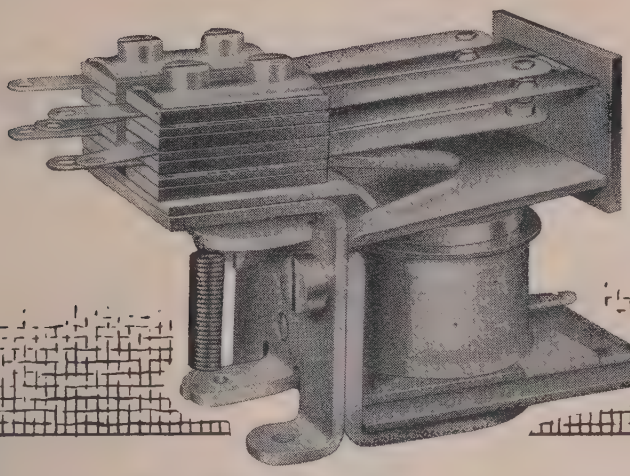
MAXIMUM OPERATING CONDITIONS (CLASS C AMPLIFIER) AT 50 Mc/s									Maximum frequency at reduced ratings (Mc/s)
Valve	Type	V_a (V)	V_{g1} (V)	I_a (mA)	I_{g1} (mA)	V_{in} (peak) (V)	P_{load} (W)	η %	
QY3-65 (CV1905)	**TETRODE**	**3000**	**-100**	**115**	**10**	**170**	**224**	**81**	**250**
TY2-125 (CV1924)	TRIODE	2500	-200	205	40	390	310	76	200
QY3-125 (CV2130)	TETRODE	3000	-150	167	6.5	300	300	75	200
QY4-250 (CV2131)	TETRODE	4000	-225	312	9	374	800	80	120

Mullard

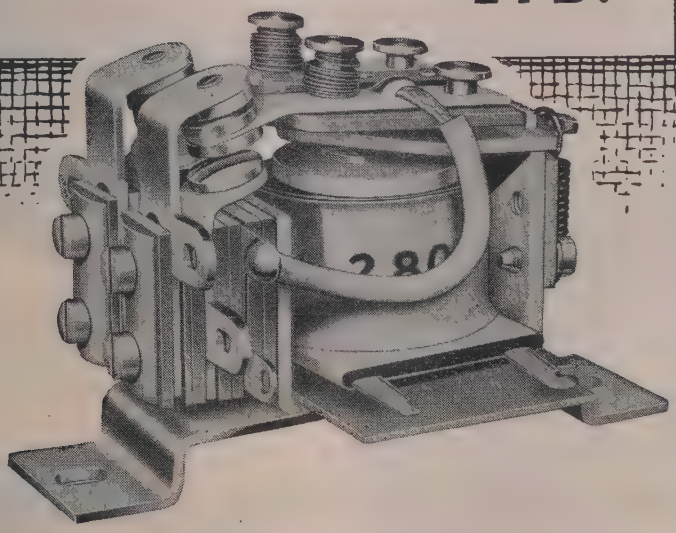
MULLARD LTD., COMMUNICATIONS & INDUSTRIAL VALVE DEPT.,
CENTURY HOUSE, SHAFTESBURY AVENUE, LONDON, WC2

MVT184

When it's **RELAYS** contact



Magnetic Devices LTD.



Upper illustration: Series 100 a.c. operated; Series 105 d.c. operated. Contact ratings up to 10 amps. continuous. Switching: one to six poles in various combinations. Overall size: $2\frac{7}{8}''$ long by $1\frac{1}{8}''$ wide by $1\frac{3}{4}''$ deep.

Lower illustration: Series 151 a.c. operated; Series 156 d.c. operated. Contact ratings up to 15 amps. continuous. Switching: double pole change-over or double pole single throw. Overall size: $2\frac{31}{32}''$ long by $1\frac{7}{8}''$ wide by $1\frac{7}{8}''$ deep.

Coils are wound for standard voltages up to 250V. A.C. and 140V. D.C. Consumption is 3 watts D.C. or 6 V.A., A.C. Coils can be supplied vacuum impregnated.

Please write for illustrated leaflet.

MAGNETIC DEVICES LIMITED

Magnetic Devices LTD.

EXNING ROAD NEWMARKET

Telephone: Newmarket 3181/2/3.

Telegrams: Magnetic Newmarket.

H.F. POWER TRANSFORMERS

RATING . . . . . UP TO 2 kW

FREQUENCY RANGE . . . . 2 Kc/s to 2 Mc/s

H.F. power transformers of outstanding efficiency are the latest addition to the Mullard range of high quality components designed around Ferroxcube magnetic cores.

Utilising the unique characteristics of Ferroxcube to the full, Mullard H.F. transformers are smaller, lighter, and less costly than transformers using alternative core materials. These advantages are particularly marked in transformers required to handle powers of up to 2kW, between the frequency range 2kc/s to 2Mc/s.

Mullard transformers are already finding wide use in applications as diverse as ultrasonic H.F. power generators and aircraft power packs operating from an aircraft's normal A.C. supply. In the latter application, the low leakage field of Ferroxcube can eliminate the need for external screening, thereby reducing the size and weight of the transformer even further.

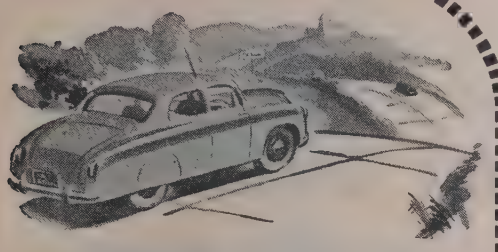
As with all Mullard high quality components, these H.F. power transformers are designed and built to engineers' individual specifications. Write now for details of the complete range of components available under this service.

Mullard

'Ticonal' permanent magnets
Magnadur ceramic magnets
Ferroxcube magnetic cores

MULLARD LTD · COMPONENTS DIVISION · CENTURY HOUSE · SHAFTESBURY AVENUE · WC2
MC 1

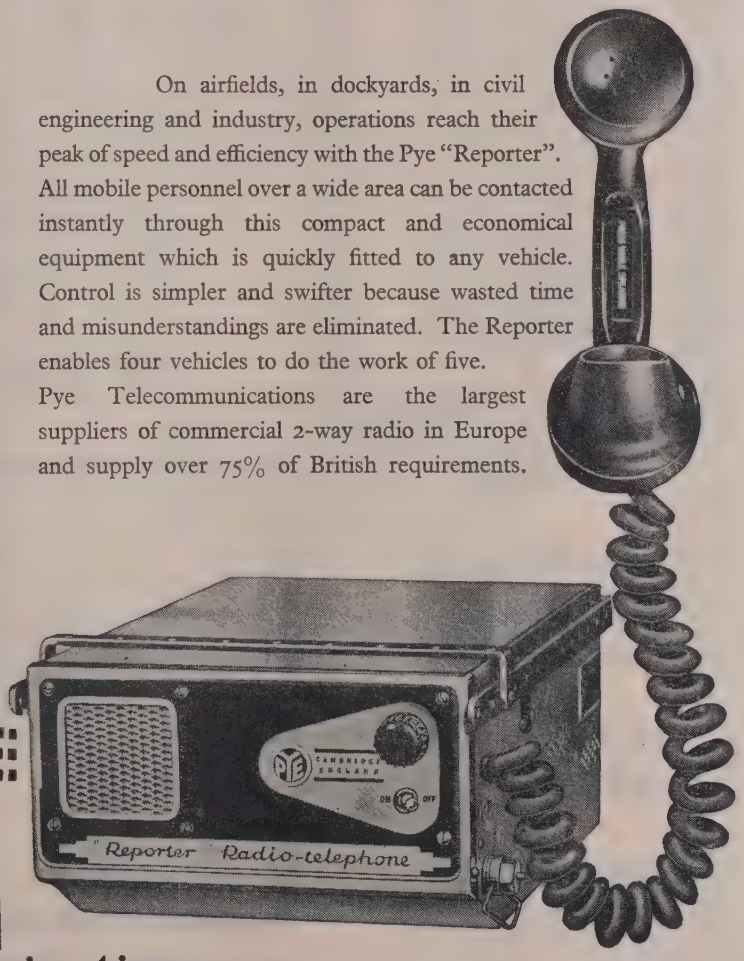
PYE REPORTER



LOW COST—SMALL SIZE
LOW BATTERY DRAIN
MOBILE RADIO TELEPHONE

On airfields, in dockyards, in civil engineering and industry, operations reach their peak of speed and efficiency with the Pye "Reporter". All mobile personnel over a wide area can be contacted instantly through this compact and economical equipment which is quickly fitted to any vehicle. Control is simpler and swifter because wasted time and misunderstandings are eliminated. The Reporter enables four vehicles to do the work of five.

Pye Telecommunications are the largest suppliers of commercial 2-way radio in Europe and supply over 75% of British requirements.



Telecommunications

CAMBRIDGE ENGLAND

Pye (New Zealand) Ltd.
Auckland C.I., New Zealand

Pye Radio & Television (Pty.) Ltd.
Johannesburg
South Africa

Pye Canada Ltd.
Ajax, Canada

Pye Limited
Mexico City

Pye-Electronic Pty., Ltd.
Melbourne, Australia

Pye Limited
Tucuman 829
Buenos Aires

Pye (Ireland), Ltd.
Dublin, Eire

Pye Corporation of America
270, Park Avenue
New York

PYE LIMITED • • CAMBRIDGE • • ENGLAND

New!

E.H.T. CAPACITOR

Type 500 H.T. oil-filled capacitor in non-metallic case

PLASTIC CASE ELIMINATES CORONA EFFECTS

HERMETICALLY SEALED

SINGLE HOLE FIXING

Terminations		$\frac{7}{16}$" x 4 B.A.		$\frac{7}{16}$" x 4 B.A.		$\frac{1}{2}$" x 2 B.A.	
Tube Size		Dia.	Len.	Dia.	Len.	Dia.	Len.
D.C. Rating 70° C. Wkg.	Test	$\frac{13}{16}$"	$2\frac{3}{8}$"	$1\frac{1}{8}$"	$2\frac{3}{8}$"	$1\frac{1}{8}$"	$3\frac{1}{8}$"
2.0 kV.	4.0 kV.	0.05 μF.		0.1 μF.		0.15 μF.	
4.0 kV.	8.0 kV.	0.01 μF.		0.02 μF.		0.03 μF.	
5.0 kV.	10.0 kV.	0.0075 μF.		0.015 μF.		0.02 μF.	
8.0 kV.	16.0 kV.	0.002 μF.		0.005 μF.		0.0075 μF.	
10.0 kV.	20.0 kV.	0.0015 μF.		0.003 μF.		0.005 μF.	
12.0 kV.	24.0 kV.	0.001 μF.		0.002 μF.		0.003 μF.	
15.0 kV.	30.0 kV.	0.0007 μF.		0.0015 μF.		0.002 μF.	
18.0 kV.	36.0 kV.	0.0005 μF.		0.001 μF.		0.0015 μF.	
20.0 kV.	40.0 kV.	0.0003 μF.		0.0007 μF.		0.001 μF.	
25.0 kV.	50.0 kV.	0.0002 μF.		0.0005 μF.		0.00075 μF.	
30.0 kV.	60.0 kV.	—		—		0.0005 μF.	

The capacitance value is a nominal maximum for a given tube size and voltage rating, capacitance tolerance ± 20%.

DUBILIER

DUBILIER CONDENSER CO. (1925) LTD., DUCON WORKS, VICTORIA RD., NORTH ACTON, W.3.
Telephone: ACOrn 2241 Telegrams: Hivoltcon, Wesphone, London.

DN150A

Call Solartron for an engineer to give you a demonstration

An Outstanding Oscilloscope

A general-purpose D.C./A.C. Oscilloscope with facilities for direct time and voltage measurement on both continuous wave and pulse phenomena.

Max. Bandwidth:	D.C. to 10 Mc/s.
Sensitivity:	1 mV/cm. to 60 V/cm.
Time-base:	0·1 μ sec/cm. to 100 m sec/cm. with expansion up to x 5.
Internal/External Sync. or Trigger.	
Dimensions:	16″ x 10″ x 23″ long.
Weight:	Approx. 70 lb.

THE SOLARTRON ELECTRONIC GROUP LTD.

RELIABILITY—UNDER OUR 12 MONTHS' GUARANTEE, COSTS HAVE NEVER EXCEEDED 0.2% OF SALES

Thames Ditton, Surrey
Telephone: EMBerbrook 5522
Cables: Solartron, Thames Ditton

The Institution is not, as a body, responsible for the opinions expressed by individual authors or speakers.
An example of the preferred form of bibliographical references will be found beneath the list of contents.

THE PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

EDITED UNDER THE SUPERINTENDENCE OF W. K. BRASHER, C.B.E., M.A., M.I.E.E., SECRETARY

VOL. 103. PART B. No. 9. MAY 1956

DISCUSSION ON 'AN INTRODUCTION TO SOME TECHNICAL FACTORS AFFECTING POINT-TO-POINT RADIOCOMMUNICATION SYSTEMS'*

Dr. D. A. Bell (*communicated*): One detail in this otherwise excellent review which is unacceptable is the statement in Section 3.1.1 that 'noise factor' is based on the idea of an ideal receiver having infinite input impedance, and that if a receiver is matched in impedance to its aerial it thereby has a noise factor of 3 dB. It is now generally recognized that all receivers are in principle *power*-operated devices, and the use of noise factor followed after the introduction by H. T. Friis† of the concept of *available power*: the noise factor is a figure of demerit of the apparatus which, through the use of available power, is based on a hypothetical matched condition and has the virtue of being unaffected by the condition of match or mismatch in which the apparatus may be used on any particular occasion. The noise factor is the ratio of the available noise power at the output of the device to the available power which would be found if the device generated solely Johnson noise. The difference between aerial–receiver matching conditions for maximum power transfer and for optimum signal/noise ratio is due to a difference in the effective temperatures of the aerial and of the receiver input circuit.

In Table 2, the wideband figures are said to be for minimum power, regardless of bandwidth. If one assumes a continuous channel, and attempts to minimize P in the formula $C = W \log (1 + P/N)$ by increasing W while N increases only as $\sqrt{W}$, how does one obtain a specific optimum bandwidth instead of a preference for bandwidth to tend to infinity? The bandwidth of 2 300 Mc/s in system No. 3 is surely too great to be fixed by practical considerations.

In Section 3.2.3 it might be noted that non-linear distortion can be subsequently corrected provided that there is no appreciable distortion of any kind in the section of the channel between the first non-linearity and a second non-linearity operating in the opposite sense. This principle is sometimes used in audio-frequency amplifiers so that two stages in cascade may give less total distortion than would be obtained from the final stage alone.

I am glad the author makes it clear in his eqn. (2) that the possible spacing of distinguishable signal levels is a function of the number of samples, $2Tf_w$. It should be emphasized that the resulting formulae such as eqns. (3) and (4) are based on probability calculations and, therefore, are useful only when the signal consists of a comparatively long code-group,‡ e.g. of the order of 100 digits or sampling ordinates.

Mr. K. L. Rao (*India: communicated*): In Section 3.3.2, the author says, 'Diversity methods include . . . (*b*) Spaced aerial diversity, in which two or three separate aerials are spaced one or more wavelengths apart, either parallel or perpendicular to the propagation path, and the strongest received signal is automatically selected. . . .' The separation of one wavelength may be adequate in the u.h.f. range, somewhat larger spacings may be needed in the v.h.f. range, and at least two wavelengths would be needed in the h.f. range. Where rhombic antennae are used in spaced diversity for long-distance reception, the side of the rhombic must be at least greater than 2λ, and consequently one would expect the separation between the aerials to be between 4λ and 10λ to have effective space diversity. Experimental data are available for simple dipoles, where it has been found that generally a spacing of 2λ would be adequate for most of the h.f. range; such data do not appear to be available for the rhombics. In practical installations one finds separation ranging between the values mentioned above.

Also, it does not always appear to be the best method to select only the strongest of the signals to obtain optimum signal/noise ratio. A small but definite amount of noise from the other channels in space diversity appears to be essential to obtain a more satisfactory signal/noise ratio in the combined output. The earliest and somewhat crude method of paralleling the automatic gain controls of the receivers would thus appear to produce better results than the later methods of electronic switching. The recently developed 'ratio squarer' method appears to hold great promise.

In Section 4.1, the author says '. . . The average depth of modulation can be increased almost to 100% by severely clipping the speech signal with amplitude limiters, a process that leads to very little loss of intelligibility. . . .' While this method has been highly successful in point-to-point systems using amplitude modulation for maintaining a high level of modulation without sideband splashing, and thus obtaining a high signal/noise ratio, it is not entirely satisfactory when one has to listen for long periods. A case in point is the programmes relayed by the Voice of America station at Manila. The clipping method, even though it scores on intelligibility, is noisy. The compandor method would appear to be more satisfactory, although it demands additional equipment.

I suppose it is generally true in all communication systems that by the reduction of working speed a better signal/noise ratio and better intelligibility are obtained. This appears to be

* LAVER, F. J. M.: Paper No. 1925 R, November, 1955 (see 102 B, p. 733).
† FRIIS, H. T.: *Proceedings of the Institute of Radio Engineers*, 1944, 32, p. 419.
‡ BELL, D. A. and DUGGAN, T. C.: *ibid.*, 1954, 42, p. 1569.

particularly true in circuits where the signal/noise ratio is poor. The same appears to be applicable to the recovery of signals in noise.

The weakest link in any communication system appears to be the receiver, with its conflicting requirements of selectivity and fidelity. The only solution appears to be to aim at obtaining the rectangular pass-band characteristics. With the advent of mechanical filters, one step appears to have been gained in this field. It may be worth while to develop circuits for obtaining the ideal pass band, as all other methods so far developed appear to be only compromises between the various circuit requirements.

Mr. F. J. M. Laver (*in reply*): Dr. Bell has drawn attention to the fact that when writing of 'noise factor' it is necessary to say in what sense the term is used, for it has been variously defined since its introduction. The definition underlying Section 3.1.1 is that quoted in C.C.I.R. Recommendation No. 94 (see Reference 12), namely: 'the noise factor is the ratio of noise power measured at the output of the receiver to the noise power which would be present at the output if the thermal noise due to the resistive component of the source impedance were the only source of noise in the system.' This agrees substantially with the definition adopted by the British Standards Institution.* These definitions are not based on the hypothesis of matched impedances and lead to noise factors that depend on the source and receiver impedances, and in particular to a noise factor of 3 dB or more for the matched condition. Dr. Bell follows Friis's definition of noise figure and claims that it is a measure characteristic of the receiver itself and unaffected by the particular impedances used on any occasion. However, it is the receiver's contribution to the total noise in some particular application that is usually of interest, and this depends not only on its noise figure, as defined by Friis, but also on the impedance-matching conditions. The alternative definition of noise factor is related to the particular source impedance of interest and shows directly by how much the receiver falls short of the ideal in its intended application. The phrase 'a maximum noise factor of 3 dB' used in Section 3.1.1 is unfortunately misleading; 'a noise factor not less than 3 dB' is intended.

The wideband figures quoted in Table 2 come from the paper by Feldman and Bennet (see Reference 18); they do not represent absolute optima—for as Dr. Bell points out none exists for fluctuation noise—but serve to illustrate the bandwidth-power exchange. Feldman and Bennet comment that very wide bandwidths might assume practical importance in waveguide systems.

It is useful to have Dr. Bell's expansion of the brief note at the beginning of Section 3.2.3, that non-linear distortion cannot easily be corrected. In general, the need to avoid distortion between the points at which non-linearity occurs and is compensated limits the use of compensation in point-to-point systems. However, the instantaneous compandor mentioned in Section 4.1 and described in References 66 and 67 is one example of the intentional introduction and compensation of non-linear distortion.

Some experimental results quoted by van Wambeck and Ross (Reference 50) suggest that for spaced-aerial diversity on frequencies between 7 and 16 Mc/s there is a rapidly diminishing advantage for aerial separations of more than 200 ft, i.e. of more than 1–2λ. These results were obtained with dipole aerials, but it cannot be generally claimed that aerials must not overlap in order to act in diversity, and so that larger aerial spacings must be used with rhombics than with dipoles. Worth-while diversity gains can still be achieved when the signals in the separate aerials have correlation coefficients of 0·5 or more. Large aerial spacings are used in practice, but they could not always be justified solely in terms of transmission gain. The design of diversity receiving systems is a complex matter which must take account of the frequency in use and the type of propagation disturbance that is being experienced, as well as the grade of service required.[A]

I agree with Mr. Rao that selection of the strongest signal is not the best method of diversity combination, but I do not agree that alternative methods derive their advantage by introducing small amounts of noise from the other channels. The advantage is obtained because even the poorer channels contain additional information about the signal; this is all lost by selection, but some is retained and used by an appropriate method of combination. A better method than selection of the strongest signal is to assess the signal/noise ratios in each of the separate diversity channels, and to combine their outputs in proportions determined by these ratios. The use of the 'ratio squarer' is a simple approximation to this process. Kaplan[B] has discussed somewhat similar techniques applied in time diversity by repeatedly sampling the received signal.

It is true that clipping introduces noise; it also changes the character of the received speech, but where intelligibility and economy are major factors it is a valuable technique. In order to obtain the full benefit of a reduction in working speed, the bandwidth of the system must also be reduced, which reduces the amount of noise and interference received and improves the signal/noise ratio. If the improvement is particularly marked when the signal/noise ratio is low, it is possibly because the signal/noise ratio passes from below to above the threshold value mentioned in Section 3.1.2. I do not agree with Mr. Rao's implied conclusion that a rectangular bandpass characteristic in the receiver would enable all other signalling techniques to be ignored.

(A) Jelonek, Z., Fitch, E., and Chalk, J. H. H.: 'Diversity Reception. Statistical Evaluation of Possible Gain', *Wireless Engineer*, 1947, **24**, p. 54.

(B) Kaplan, E. L.: 'Signal-Detection Studies with Application', *Bell System Technical Journal*, 1955, **34**, p. 403.

* British Standard 2065: 1954. Glossary of Terms for Characteristics of Radio Receivers. No. 202.

621.396.71 : 621.396.61(425)

Paper No. 1903 R
Oct. 1955

THE NEW HIGH-FREQUENCY TRANSMITTING STATION AT RUGBY

By Captain C. F. BOOTH, O.B.E., and B. N. MacLARTY, O.B.E., Members.

(*The paper was first received* 30*th March, and in revised form* 25*th May*, 1955. *It was published in October*, 1955, *and was read before* THE INSTITUTION 3*rd November*, 1955, *and the* RUGBY SUB-CENTRE 14*th February*, 1956.)

SUMMARY

A new radio transmitting station has recently been built by the Post Office on a 700-acre site adjacent to the original Rugby station. The 28 transmitters installed are suitable for independent-sideband operation in the frequency range 4–27·5 Mc/s, provide 30 kW peak envelope power and can carry both multi-channel telephone and multi-channel telegraph signals. Upwards of 70 aerials (mainly rhombics) will be provided and will cover transmissions beamed on all the main traffic routes.

In the design of the station attention has been focused particularly on flexibility, whereby the maximum use can be made of the plant with the minimum staff. This has been achieved by careful attention to the design of the building; by the maximum standardization of equipment; by the segregation of low- and high-power equipment; by the provision of remotely controlled frequency-changing and aerial-switching; and by the inclusion of a central control position from which the station is operated. Staff at this position can start and stop any transmitter, switch to any one of six predetermined frequencies and to the required aerial, fully monitor its performance and quickly locate a fault.

The new station was brought into service in October, 1954, and formally opened on the 28th July, 1955; experience of its operation has fully justified the basic design considerations employed in its planning.

(1) INTRODUCTION

The telecommunication network connecting the United Kingdom with other parts of the world is a closely-knit combination of submarine-cable and radio circuits. Those cables for which the United Kingdom is responsible comprise the Post Office multi-channel telephone cables to the Continent and the world-wide submarine-telegraph-cable system of Cable and Wireless, Ltd. The radio circuits, the United Kingdom terminals of which are operated by the Post Office, provide public telephone services, and public, private and multi-destination telegraph services to fixed stations abroad and to ships at sea. It is difficult to show the multi-destination and ship services on a map, but the extent of the other radio services is illustrated in Fig. 1. In all, some 23 U.K. radio stations participate in the oversea fixed and mobile services, using radio frequencies extending from 16 to 27 000 kc/s, but most of the long-distance point-to-point services are operated in the h.f. band (3–30 Mc/s) using transmitting stations at Bodmin, Criggion, Dorchester, Leafield, Ongar and Rugby, and receiving stations at Baldock, Bearley, Brentwood, Cooling and Somerton.

Despite expansion of the existing stations and the provision of a new transmitting station at Criggion[1] during the war, it became apparent immediately after the war that within a few years the expansion of telephone and telegraph traffic would necessitate the provision of large new transmitting and receiving stations. The planning of these stations was initiated and sites were selected adjacent to the existing Rugby station and at Bearley near Stratford-on-Avon, for the transmitting and receiving stations respectively. The new Rugby station was brought into service in October, 1954, and the Bearley station

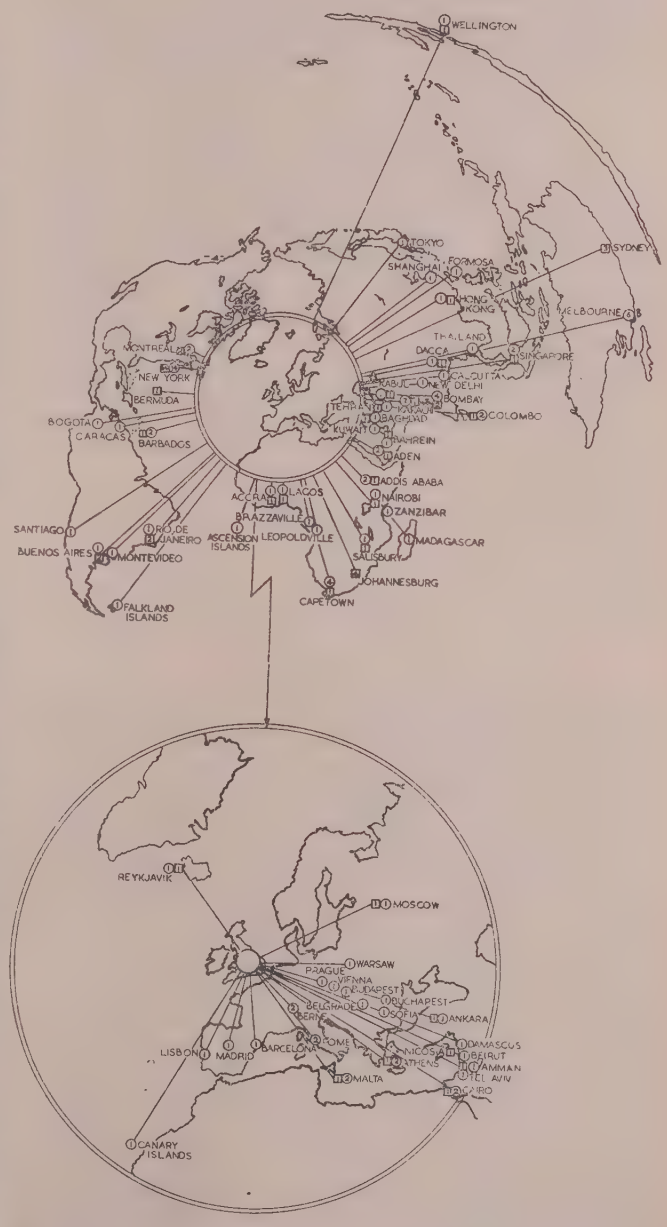


Fig. 1.—Oversea point-to-point radio services operated in the U.K. by the Post Office.

□ Number of telephone channels.
○ Number of telegraph channels.

Capt. Booth is in the Post Office Engineering Department.
Mr. MacLarty is with Marconi's Wireless Telegraph Co., Ltd.

in February, 1953. These two new stations, with the radio-telephony terminal at Brent, represent the first major construction projects for the oversea radio services carried out by the Post Office since the war. They incorporate the most recent developments, and it is the purpose of the present paper to describe the design of the new Rugby station and its equipment.

(2) DEVELOPMENT OF THE EARLIER RUGBY STATION

Very soon after its opening in 1926, the original Rugby radio station[2,3] became well known to radio operators and communication engineers in many lands through its world-wide telegraph transmissions on 16 kc/s—the only transmissions made from the station at that time. The aerial for this 350 kW transmitter, mounted on 12 lattice masts 820 ft high, is an impressive sight as the station is approached. In 1926, a single-sideband (s.s.b.) 50 kW transmitter[4] operating at 60 kc/s was installed and used in 1927 for the first public long-distance radio-telephone service (London–New York) to be operated anywhere. Following the spectacular success of high frequencies for long-distance radio communication, in 1927 the first h.f. transmitters were installed at Rugby. To-day (1955) the earlier station contains 18 h.f. transmitters, most of which are used for independent-sideband (i.s.b.)* operation[5] and can provide up to four telephone channels per transmitter. In addition, four transmitters have been installed for standard-frequency transmissions[6] and one for telegraph traffic at 64 kc/s. It will be appreciated from this brief survey that the earlier Rugby station has developed with the art and, despite the advent of the new station, will remain a very important part of the oversea radio network.

(3) BASIC DESIGN CHARACTERISTICS OF THE NEW RUGBY STATION

The new Rugby station is required to transmit telephone and telegraph traffic to many parts of the world, and has been designed to accommodate 28 h.f. transmitters and upwards of 70 aerials.

(3.1) Transmitters

For telephone operation the merits of the i.s.b. system are so outstanding that for some time the system has been standard in the Post Office, and it was therefore adopted for the telephone transmitters at the new station. The position with regard to the equipment for the telegraph services was not so clear. A variety of systems, both single- and multi-channel, was being used for public traffic, and, in addition, important multi-destination Press services were being operated on various different single-channel systems. Thus there were potential requirements for single- and multi-channel telegraph systems using frequency- or time-division multiplex, or a combination of the two, in association with frequency-shift or on-off keying. There was also a need for facsimile transmissions.

Despite this wide variety of service requirements it was considered essential to standardize as much as possible of the transmitting equipment, in order to achieve the economic and operational advantages that such a policy yields. At the same time it was necessary to provide flexibility, so that future changes in the demand for the different classes of service might easily be met. By dividing a complete transmitter into two parts, a low-power drive unit and a high-power transmitter-amplifier, it has been possible to make the latter (the larger and more costly part of the transmitter) suitable for all the various classes of service. The low-power fixed-frequency drive units are of two types, one catering primarily for multi-channel i.s.b. telephony, and the other for single- or multi-channel telegraphy, including facsimile, using frequency-shift or on-off keying. Common levels and mid-band frequencies are used for the output signals of both types of drive unit, so that the transmitter-amplifiers can be driven from either without readjustment. The transmitter-amplifier accepts the low-power signals from the drive, converts them to their final frequency and raises them to high power.

Labour charges bulk so largely in the annual cost of running a transmitting station that it is of the highest importance to keep the cost of staffing to a minimum. Much use has therefore been made of remote-control and remote-monitoring facilities. The arrangements are such that any of the 28 transmitters can be switched to any one of its assigned frequencies, connected to the appropriate aerial, monitored during operation and switched off, all by an operator at a central control position; and it is necessary for staff to work on individual transmitters only when maintenance is required or new frequencies are assigned. To facilitate maintenance the various types of low-power equipment, e.g. the drive units and the fixed-frequency oscillators, are grouped together with appropriate test equipment, and all are concentrated in one room from which connections are made to the transmitter-amplifiers.

Initially, consideration was given to grouping the transmitter-amplifiers in several unattended buildings relatively near to the aerials they would normally serve. However, the advantage of reducing the aerial-feeder losses by such an arrangement was considered to be less than the countervailing advantages of the improved flexibility and easier maintenance that would result if all the transmitters were concentrated in a central building. Accordingly, the 28 transmitter-amplifiers are housed in three wings converging on a central hall containing the central control position.

To determine the powers of the transmitters for a station that is to cater for future development is a difficult problem involving consideration of many factors. For example, the receiving equipments used for the main public telegraph and telephone services are more elaborate than those used on the multi-destination services, and the economic balance between expenditure on transmitting and receiving equipments is different in the two cases. The outstanding importance of some of the public traffic routes, and the rapidly expanding use on those that are busiest of multi-channel systems for telephony and telegraphy and perhaps for both simultaneously, are factors that tend to increase the power requirements of the transmitters used on them. In addition, the variability of radio-propagation conditions from day to day, month to month and year to year, and the changing pattern of services and techniques, introduce major uncertainties which usually make it impossible to predict with accuracy the transmitter power needed to provide a given grade of service. Nevertheless, long experience suggests that, for the more important and the more difficult routes, telephone transmitter powers of 30 kW peak envelope power (p.e.p.) and telegraph transmitters of 20 kW peak power are desirable, and since transmitters of lower power were available at existing stations for use over the easier routes, it was decided that all the transmitters at the new station should be of 30 kW p.e.p. on independent-sideband telephony and 20 kW on frequency-shift telegraphy.

High-power transmitters can be designed with either water or air cooling, the relative advantages and disadvantages of the two methods depending to some extent on operational and climatic conditions. For Rugby it was felt that air-cooling was preferable, provided that its noise was not troublesome, and it was adopted after tests to confirm this.

It was decided that the overall performance of the transmitters in respect of frequency stability, spurious emission, etc., should

* I.S.B. operation is a form of multi-channel working in which the radiated signal comprises a reduced carrier wave and two or more sidebands conveying independent signals.

be in advance of internationally agreed standards in anticipation of possible upward revision of these standards and in the interests of other users of the ether.

(3.2) Aerials

The guiding principle in the provision of aerials has been to provide the highest gain compatible with reasonable cost and site area, and in the experience of the Post Office this can best be achieved with rhombic aerials. Because of their simplicity, good all-round performance and relatively wide frequency coverage they compare favourably with other types. Moreover, their aperiodic property simplifies aerial-switching requirements, enables them to be designed and erected regardless of particular frequency assignments and avoids modification for subsequent frequency adjustments should the initial frequency be unsuitable. With the increasing difficulty in obtaining satisfactory assignments, this is of considerable importance. In view of these considerations it was decided to use rhombic aerials for all except the New Zealand service—a route for which the best direction of transmission varies over a range greater than can readily be covered by a rhombic—and for this a stacked array of dipoles is used.

(3.3) Transmitter-Aerial Connections

Ideally it is desirable to provide switching facilities whereby any transmitter can be connected to any aerial, but to do this for 28 transmitters and 70 or more aerials would require an extremely complex switching system.

The other extreme would be a one-to-one allocation of transmitters to services, but even if such an allocation could be made and left unchanged for long periods, it would still be necessary to cater for daily frequency changes on each service, which would sometimes involve aerial changes. In addition, some services are operated on relatively short daily schedules, and one transmitter may be used successively for several services on different routes. At a large station it is also desirable to have a still higher degree of flexibility to cater for changes in operational requirements and to facilitate maintenance. It is therefore clear that a one-to-one allocation would be both uneconomic and impracticable. The switching arrangement which has been adopted lies between the extreme solutions: both the transmitters and aerials have been divided into two equal-sized groups with full flexibility within each group and some cross-connections between groups.

The leading-in of upwards of 70 transmission lines to the building, and their extension to the transmitters via the switches, involves bringing a large number of feeders into a relatively small space. For this purpose cables have advantages over open-wire feeders in respect of safety, ease of switching and freedom from crosstalk. On the other hand, for the relatively long transmission lines from the building to the aerials, cables are much more expensive than open-wire lines of equal loss. Coaxial cables have been used within the building and open-wire feeders outside, the interconnections being made by aperiodic matching lines.

(4) SITE AND BUILDING

(4.1) Site

The main requirements for a large transmitting-station site are an adequate area of flat open land of good electrical conductivity extending for at least 500 yd beyond the site proper, with no high ground subtending an angle of more than about 2° at the boundary; access to a reliable power supply and to the trunk cable network; and adequate housing and amenities. It was found that these requirements could best be met by acquiring further ground near the existing Rugby site. Moreover, such an arrangement would facilitate joint administration of the two stations, with consequent economies in staff and in facilities, such as emergency power, land-lines and transport. As a result, the adjacent site was selected and some 700 acres of mixed arable and pasture land was purchased. Fig. 2 shows the site plan. Tenancies have been arranged with present occupiers, and in laying out and installing the external plant every effort has been made to keep interference with the agricultural interests to an absolute minimum.

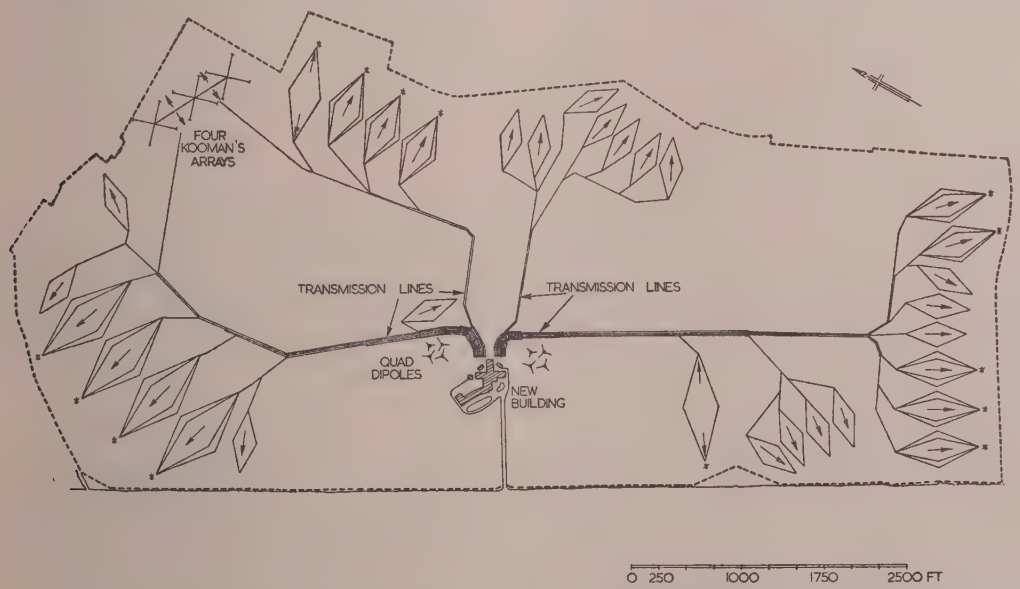


Fig. 2.—Site plan of new Rugby radio station.
* Denotes two rhombics on same set of masts.

(4.2) Building

The station building has been designed by the Chief Architect's Division of the Ministry of Works, and since the architectural

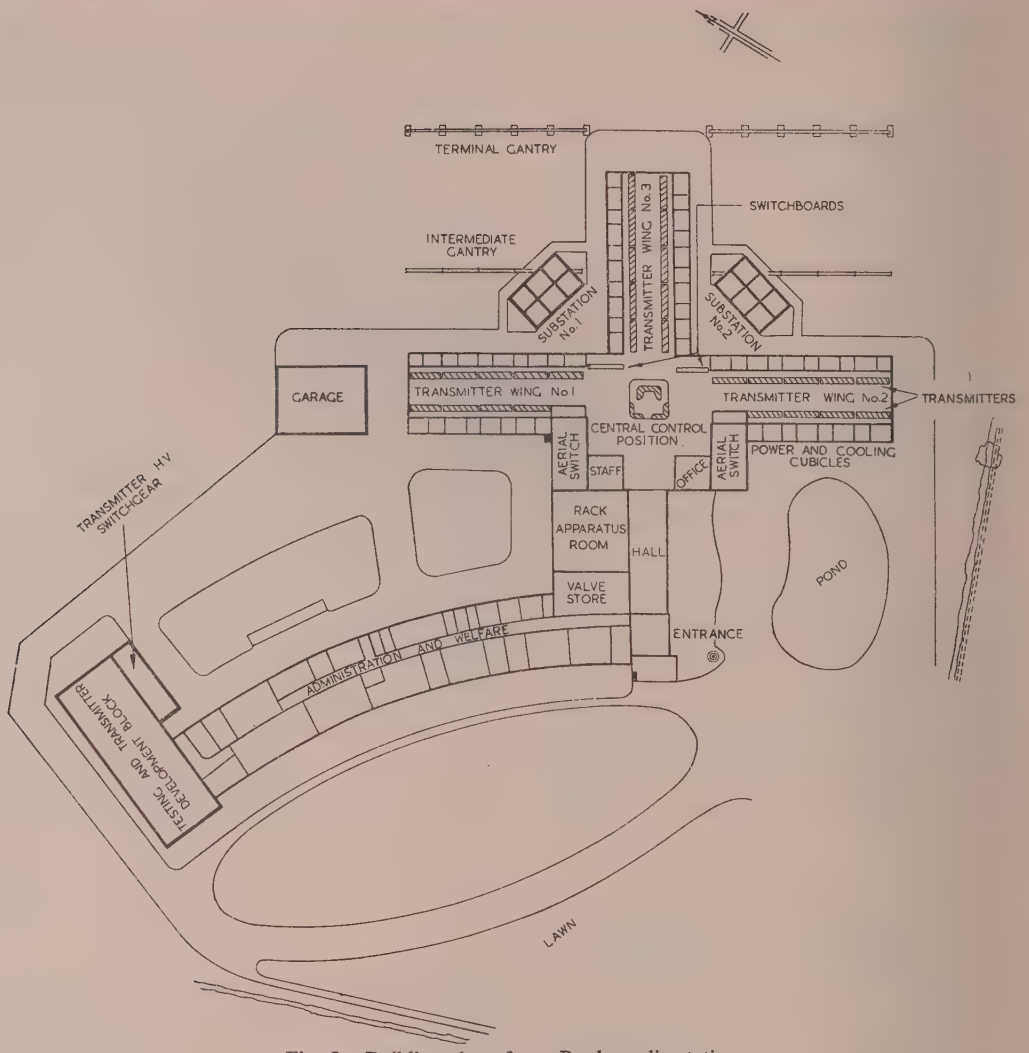


Fig. 3.—Building plan of new Rugby radio station.

and civil-engineering aspects will be described elsewere,[11] it is not proposed to discuss them in any detail here. The building, which is of single-storey construction, is situated approximately in the middle of the site and is divided functionally, the layout and key plan being shown in Fig. 3. The curved wing facing the road contains the administrative and welfare accommodation, including the workshop, stores and canteen, and terminates at the northern end in a testing and transmitter-development block. On the south side facing the entrance there is a pond which serves the double function of providing a static water supply in the event of fire and of making the approach to the station more attractive.

Each of the three wings housing the 28 high-power transmitter-amplifiers and their associated power-supply equipments is 105 ft long and 30 ft wide, and is designed to accommodate two rows of five transmitters. In each of two wings, however, one transmitter position is used as a local valve store. Each wing is flanked on either side by cubicles, two being assigned to each transmitter. One which is fireproof contains the oil-filled equipment (power transformers, smoothing chokes and capacitors) and is accessible only through an external door; the other houses the fan and ducting for the air-cooling system. The area at the intersection of the three wings contains a glazed enclosure housing the central control position (c.c.p.), which is raised some 2 ft above the general floor-level and gives an uninterrupted view along the three wings. The c.c.p. constitutes the nerve centre of the station, where all the operational and supervisory functions are performed.

The main corridor connecting the operational and administrative sections of the building is flanked on one side by the rack-apparatus room, 48 ft × 44 ft, in which all the telephone- and telegraph-drive equipment, the fixed-frequency oscillators, the monitoring and line equipment, and the necessary testing equipment are concentrated. Since the power dissipated may be as high as 50 kW, forced ventilation by filtered air is provided via overhead ducts.

Adjoining the central area of the transmitter wings are two rooms containing the switches through which the transmitter outputs are connected to the aerials. Behind the transmitter wings, i.e. on the side remote from the main entrance, there are two separate substation buildings containing the transformers and switchgear terminating the incoming power feeders; on one side there is a separate garage block.

The hot-water central-heating system is capable of heating

the station unaided, but, to save fuel, arrangements have been made to use the heat from the transmitters by controlled circulation of the cooling air.

Fluorescent lighting is provided throughout the operational sections of the building, and wall-mounted reflectors are used to achieve general indirect lighting from walls and ceiling.

(5) INTERNAL PLANT

Excluding the power equipment, which is considered separately, the internal plant comprises the following major items: the land-line equipment; the transmitter drive units; the fixed-frequency oscillators; the monitoring equipment, all of which is installed in the rack-apparatus room; the transmitter-amplifiers; the central control position; and the aerial switches, which are installed in or near the transmitter wings.

(5.1) Line Terminal Plant

Land-lines in two different main-cable routes connect the telephone and telegraph terminal offices in London to the new station. There are 18 telephone channels transmitting frequencies up to 6 kc/s, three multi-channel voice-frequency telegraph systems and a number of order wires and monitor channels for telephony and telegraphy. Six of the telegraph channels are capable of operating at speeds up to 280 bauds, and 12 have a maximum speed of 140 bauds. There are also two tie-cables between the old and new stations, one to give greater flexibility in the use of land-line circuits by providing for cross-connections between the stations, and the other for miscellaneous local services. All the land-line circuits to the new station are terminated in the rack-apparatus room. First they are connected to equalizing and amplifying equipment, then through a jumper field for long-term flexibility, and finally through a patching field for testing purposes and short-term flexibility. From this point the telephone channels pass directly to the telephone-drive units, and the 3 kc/s channels terminate in the telegraph-terminal equipment from which individual telegraph channels pass to the telegraph drive units.

(5.2) Drive Units

Before discussing the drive units themselves in detail it will be helpful to consider in general terms the functions they are required to perform. As indicated in Section 3.1, there are two types of drive unit, one for i.s.b. telephony and the other for telegraphy, and either is suitable for direct connection to a transmitter-amplifier. The drive units (Fig. 4) transform appro-

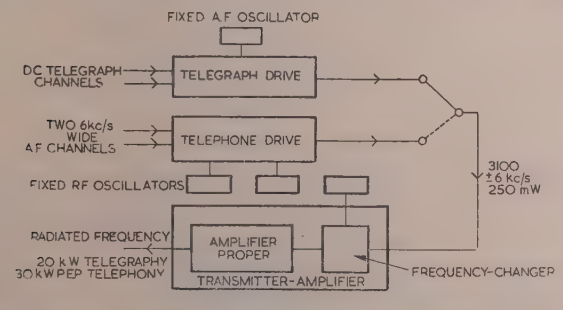


Fig. 4.—Block schematic showing relationship between drive units and other equipment.

priate land-line signals into a suitable form and translate them into a frequency band centred on a low-level pilot carrier of 3·1 Mc/s, the peak power at this point being 250 mW. The signal is then passed over a 75-ohm cable to the transmitter-amplifier, where it is translated to its final frequency before being applied to the amplifier proper. A number of the oscillators needed in these processes are not incorporated in the drive units or the transmitter-amplifiers but are concentrated in a part of the rack-apparatus room.

The telephone drives provide two 6 kc/s channels, each capable of accommodating two 3 kc/s speech channels or any desired combination of telephone and telegraph signals of appropriate overall bandwidth. The telegraph drives can provide a number of different types of telegraph signal: single-channel A_1 or A_2 signals, single-channel or twinplex F_1 signals, and F_4 signals. There are 20 telephone and 20 telegraph drive units, either type being available via switches for 12 of the transmitter-amplifiers, while telephone drives only are available for a further eight transmitter-amplifiers and telegraph drives only for the remaining eight. With such an arrangement, therefore, as many as 20 of the 28 transmitters can be operated simultaneously with either telegraph or telephone drives—a sufficient degree of flexibility for all expected requirements.

(5.2.1) Telephone Drive Units.

In the design of the telephone drive units particular attention was paid to: ease of maintenance, by providing highly stable performance and automatic monitoring; minimum spectrum occupancy, by full use of the 6 kc/s bands flanking the pilot carrier with high rejection outside; and assisting automatic frequency control (a.f.c.) at the distant receiver, by providing a clean pilot carrier.

The two audio input channels, which can extend from 100 c/s to 6 kc/s, pass through separate chains each comprising an audio filter, a line-amplifier, a limiter, a volume indicator, a balanced modulator, a sideband filter and an intermediate-frequency amplifier, as shown in Fig. 5. The audio filter rejects any 50 c/s hum accompanying the signals and prevents the production of sidebands so close to the pilot carrier that they could disturb the distant receiver a.f.c., while high-frequency cut-off reduces out-of-band radiation.

The negative-feedback line-amplifier includes a neon peak-limiter which also shows when the signal level exceeds the normal "line-up" level. The limited output cannot overload subsequent stages, with a consequent loss of intelligibility in the other sideband due to intermodulation, although the signals undergoing limiting must, of course, suffer distortion.

Each audio output is monitored by a volume indicator before passing to the first balanced modulator, where it is translated to the region of 100 kc/s. Two crystal filters select respectively the upper sideband for sideband A and the lower sideband for sideband B; these are amplified separately and then combined in a hybrid circuit. The combined signal is fed through a further hybrid circuit to the monitor receiver and to the main path, where a 100 kc/s stop filter eliminates any residual out-of-balance carrier component from the modulators. A controlled low-level 100 kc/s pilot carrier is next combined with the two independent sidebands, and the resulting signal is passed to the second modulator, where it is translated to a band centred on a 3·1 Mc/s pilot carrier. The signal is then raised to a peak sideband level of 250 mW, which is monitored by a peak valve-voltmeter. The monitor receiver (Fig. 6) produces audio-frequency outputs for intermodulation and other tests, measures the pilot-carrier level and operates an alarm when this level changes by more than a predetermined amount.

On occasion there may be a need to radiate double-sideband (d.s.b.) signals, e.g. when a demand is made for a channel to a receiving station not fitted with an s.s.b. receiver, and to cater for this contingency a d.s.b. signal can be generated in the B-channel equipment. The signal follows the normal paths to the

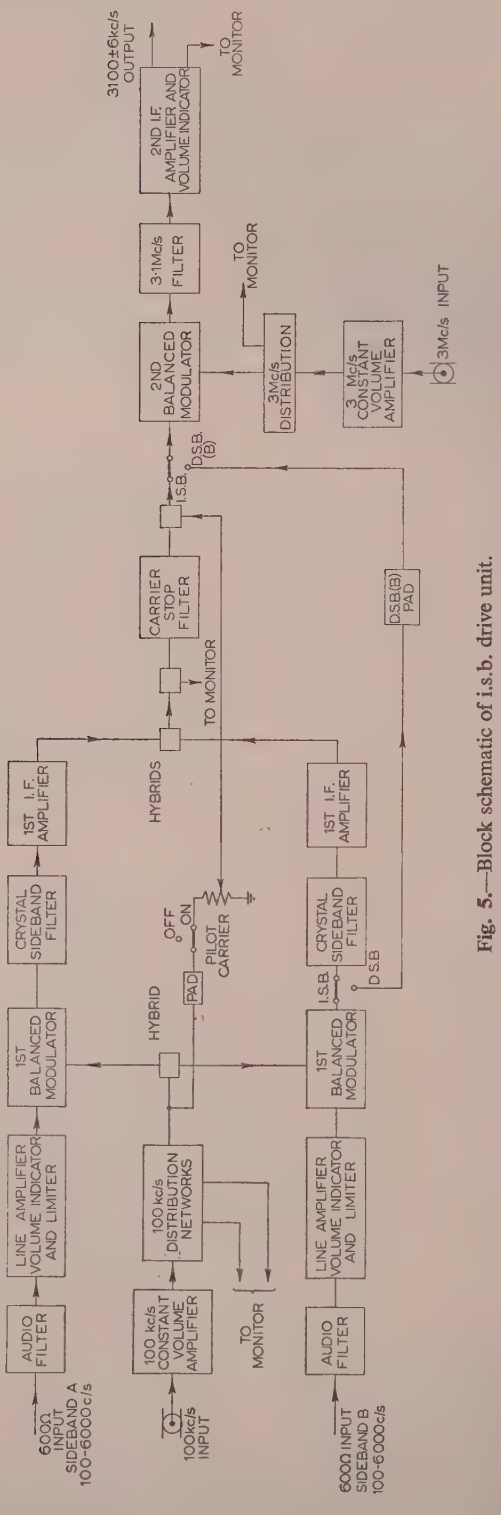


Fig. 5.—Block schematic of i.s.b. drive unit.

first balanced-modulator, which is, however, unbalanced to give a d.s.b. signal with a carrier of the correct amplitude and phase. This signal is then fed to the second balanced modulator for translation to the mid-band frequency of 3·1 Mc/s. The equipment which normally adds the other sideband and the pilot carrier is by-passed. From the second modulator the d.s.b. signal follows the normal path.

The output of the 100 kc/s oscillator is filtered and fed to a single-stage constant-volume amplifier to stabilize the levels of the carrier supply to the first modulators and of the re-inserted pilot carrier. The level of the 3 Mc/s input to the second balanced modulator is similarly stabilized.

Gain stability is ensured by feeding all stages of the drive equipment from a 280-volt stabilized supply, with a hum output of less than 1 mV. The anodes of the beam tetrodes in the final stage of the second i.f. amplifier are fed directly from an unstabilized, but well-smoothed, 400-volt line. Special precautions, including the use of Mumetal screens in the a.f. stages, are taken to reduce hum.

The equipment is housed in a 7 ft 6 in aluminium-alloy cabinet, and all units containing valves are mounted on runners, which permit them to be withdrawn to the front without interrupting operation. Passive circuits are mounted on plates accessible through the rear door. Level controls for the pilot carrier and the test tones, a feed meter and switch, and a.f. and r.f. U-links for testing and cross-patching, are mounted on the front.

The performance of the drive unit is summarized below.

A.F. Inputs.—+10 to −10 dB relative to 1 mW from 600-ohm lines.

Signal Output.—250 mW p.e.p. into 75 ohms; the carrier level is normally −26 dB, but is adjustable between 0 and −32 dB relative to 250 mW.

Non-linear distortion.—Second- and third-order intermodulation products are at least 50 dB below the level of either tone in the standard 2-tone test.

Frequency response.—The overall frequency response for the two sidebands as is shown in Fig. 7.

Gain stability.—The pilot carrier level is stable to within 0·5 dB between 0 and 55° C for 2% mains-voltage change.

Noise level.—Better than −66 dB relative to 250 mW.

Spurious and harmonic outputs.—All better than −60 dB relative to 250 mW.

Double-sideband operation.—250 mW p.e.p. at 100% modulation with less than 2% distortion.

(5.2.2) Telegraph Drive Units.

The telegraph drive unit provides a 3·1 Mc/s keyed signal for any one of the following types of emission: 2-channel frequency-shift keying (f.s.k.) for telegraphy; single-channel frequency-shift keying for telegraphy and facsimile; single-channel on-off keying of continuous waves or modulated continuous waves (modulated from an external source); and single-channel on-off keying of frequency-modulated continuous waves. The equipment is contained on two 19 in panels consisting of a keying unit and a monitor unit.

The keying unit (Fig. 8) comprises a 3·1 Mc/s crystal oscillator coupled to a reactor-valve stage which is controlled by keying signals from the incoming lines A and B; after amplification the keyed signal is fed to the output and monitor sockets. A switch enables various forms of mechanical or electrical keying to be employed; keying potentials for both channels are provided by the unit when mechanical keying is used. For test purposes an internal electrical keying position provides reversals at mains frequency, the signal on channel A being suitably phased relative to that on channel B to produce a stepped waveform with 2-channel f.s.k. operation.

For single-channel f.s.k. working the incoming d.c. keying signals can be connected to a signal inverter stage. This is used to correct the frequency inversion that takes place in a subsequent

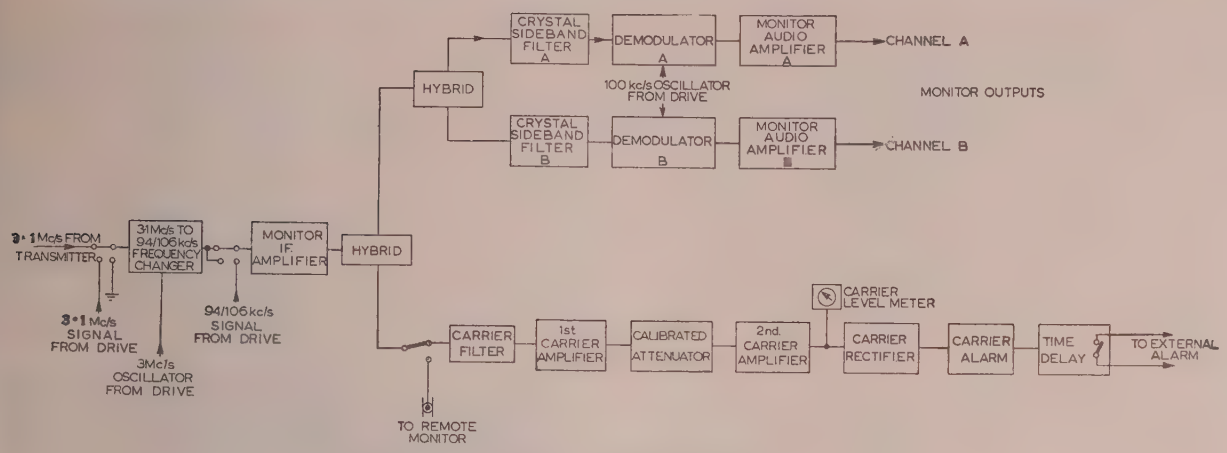


Fig. 6.—Block schematic of monitoring circuit i.s.b. drive unit.

mixer unit when the radiated frequency is below 10 Mc/s, and feeds a trigger circuit which provides an output of constant amplitude, irrespective of fluctuations in the amplitude of the keying signals. For 2-channel f.s.k. operation the trigger output is taken through d.c. limiting amplifiers which select one of four distinct bias potentials to be passed to the valve reactor.

With c.w. keying the trigger output is curbed and fed to the control grid of a 3·1 Mc/s buffer-amplifier valve, a static frequency-correcting bias being applied to the reactor valve to ensure that the output frequency is exactly 3·1 Mc/s. For m.c.w. operation the c.w. signals are modulated by a tone applied to the suppressor grid of the 3·1 Mc/s buffer amplifier from an external source.

For frequency-shift working the various d.c. keying potentials derived as described above are applied to a calibrated potential-divider that controls the grid bias of the reactor valve, and so the amount of frequency shift. Tappings on a subsequent resistor chain enable the total frequency shift to be divided by two or four when the associated transmitter-amplifier uses frequency doublers or quadruplers. The reactor valve controls the frequency of the crystal oscillator by pulling it away from the condition of crystal series resonance. The temperature-controlled crystal has a resonant frequency of 3 099·4 kc/s, and is coupled to the maintaining valve by a quarter-wave network.

With 2-channel f.s.k. operation, each channel may use one of two frequencies for the space condition and one of two other frequencies for the mark condition, and two codes are in use as shown in Table 1. Code 1 is normally used since it gives a better performance at the receiver. The lowest frequency, f_1, is the series resonant frequency of the crystal; the other three frequencies, f_2, f_3 and f_4, are respectively 400, 800 and 1 200 c/s higher than f_1.

Table 1

CODES FOR 2-CHANNEL F.S.K. OPERATION

Frequency	Code 1		Code 2	
	Channel A	Channel B	Channel A	Channel B
f_1	Space	Mark	Space	Space
f_2	Space	Space	Space	Mark
f_3	Mark	Space	Mark	Space
f_4	Mark	Mark	Mark	Mark

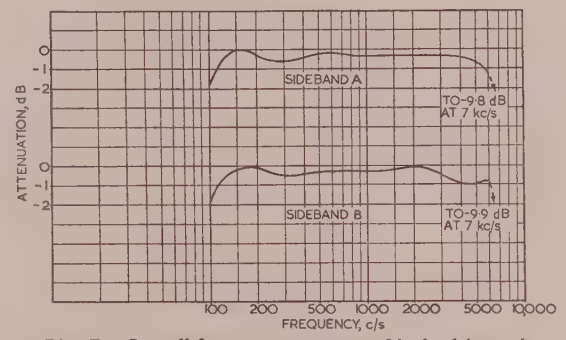


Fig. 7.—Overall frequency response of i.s.b. drive unit.

The monitor unit (Fig. 8) provides for checking the frequency shifts, either aurally or on a built-in meter, and facilitates examination of the transmitted f.s.k. waveform by means of an external cathode-ray oscillograph.

The performance of the telegraph drive units is summarized below.

Electrical key potential.—±5 volts minimum into 2 kilohms.
Signal output.—250 mW into 75 ohms.
Harmonic outputs.—Second harmonic at least 58 dB below the fundamental; third harmonic at least 65 dB below the fundamental.
Unwanted amplitude modulation on f.s.k. operation.—Less than 2% for 1 200 c/s shift.
Single-channel frequency shift.—Adjustable from 50 to 1 200 c/s.
Frequency stability.—Variations with temperature of the frequencies f_1–f_4 as shown in Fig. 9; in normal operation a stability within a few parts in 10^6 is obtained.

(5.3) Fixed-Frequency Oscillators

The internationally agreed frequency tolerance for fixed stations of power exceeding 500 watts operating in the range 3–27·5 Mc/s, is ±30 parts in 10^6 and care has been taken in the design of the fixed-frequency oscillators[7,8] to ensure that this is fully met.

The 100 and 3 000 kc/s signals and the a.f. test tones required for the telephone drive units are provided by frequency generators each having 28 outputs of 100 mW with at least 80 dB crosstalk attenuation between any pair of outputs. The 100 and 3 000 kc/s frequency generators are normally driven by a 100 kc/s signal from a frequency standard which is kept within ±2 parts in 10^8 of its nominal frequency. Standby 100 kc/s oscillator equipment of accuracy better than one part in 10^6 is provided, and changeover is effected automatically when the main signal fails.

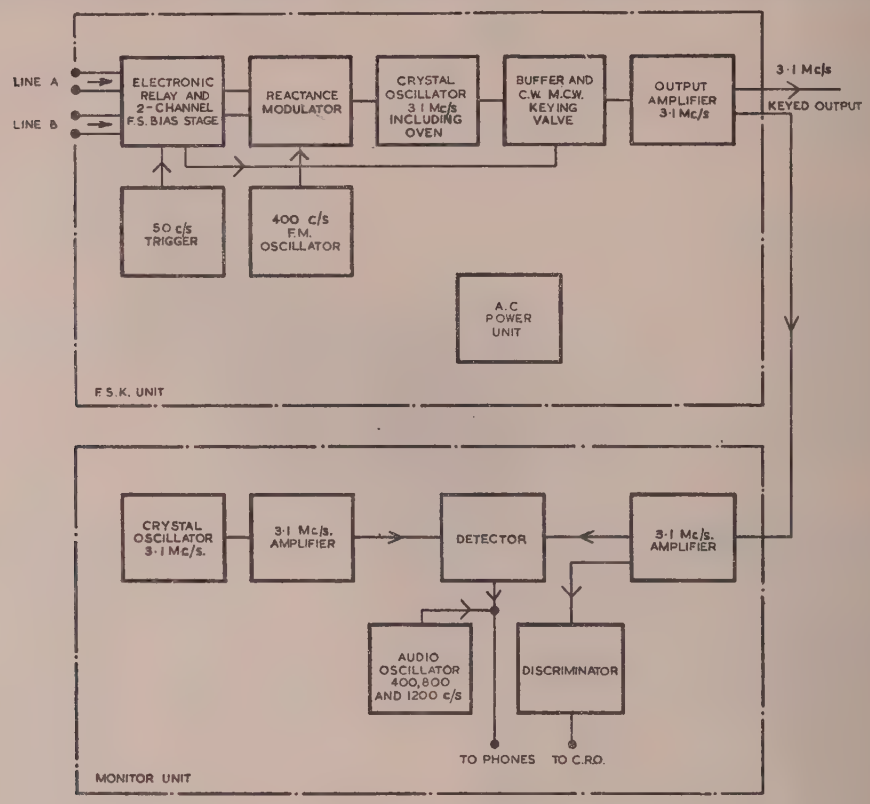


Fig. 8.—Block schematic of keying and monitor units.

The arrangements for supplying a.f. tones to the drive units for test purposes are generally similar to those for the common frequencies just described, the signals being generated by low-frequency crystal oscillators. In this case, however, purity of waveform is important, and by careful design the total harmonic content has been reduced to −58 dB relative to the fundamental.

The oscillator signals for the final frequency-changers are derived from crystal oscillators via frequency-multipliers; the frequencies generated range from 3·4 to 7 Mc/s and the output power is 100 mW. The quartz plates used have a frequency/temperature coefficient of less than one part in 10^6 per deg C, temperature control is unnecessary, and the long-term stability is better than 15 parts in 10^6 over normal indoor-temperature ranges in this country. The oscillators are simple and small, and are rack-mounted in banks of six per transmitter. In view of the great technical and operational importance of close frequency control, a frequency-measuring set is provided on which the frequency of any of the oscillators can quickly be determined by the station staff.

(5.4) Monitoring Equipment

Automatic signal monitors are provided for checking the operation of the transmitters on telephone and telegraph signals; in these the signals arriving at the station over the land-lines are compared in terms of their envelope for telephony, and element by element for telegraphy, with those derived from the outputs of the associated transmitters. The telephone monitors compare the envelopes of the land-line and transmitter-output signals in terms of the charges developed on two *RC* combinations, and each time the difference exceeds a predetermined value the event

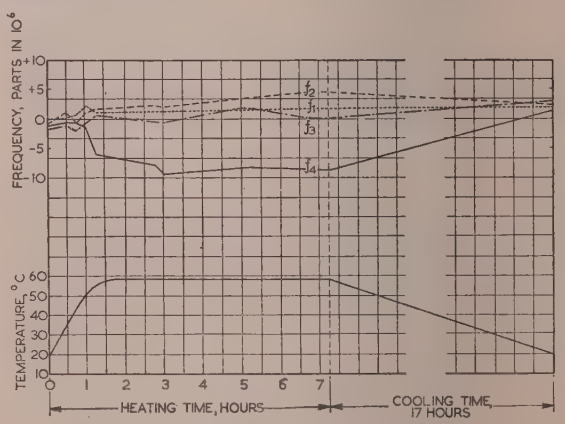


Fig. 9.—Frequency stability of telegraph drive unit.

is recorded on a simple counter; when more than four such events are recorded in half a minute, an alarm is given. The telegraph monitors in effect measure the distortion introduced in each signal element, and when this exceeds a certain value register the occurrence on counters, an alarm being given, as before, if more than four such registrations occur in half a minute. When an alarm is given by a telegraph monitor the signals arriving over the land-line are recorded, thus defining the point at which the transmission is considered to have broken down and from which retransmission is necessary.

Further automatic supervision is provided by carrier-level monitors in the transmitter output circuits, and by reflectometers

for checking the input impedance of the aerial transmission lines. If the response of either of these devices falls outside predetermined limits an alarm is given.

(5.5) Transmitter-Amplifiers

The complete transmitter-amplifier is 20 ft long, 3 ft 9 in wide and 7 ft high; it is contained in an enclosure built of pressed aluminium panels, with a number of internal free-standing iron frameworks mounting power supplies and control circuits, and with aluminium cabinets mounting the higher-power r.f. stages. A view of one of the transmitter wings showing the c.c.p. is given in Fig. 10. The equipment comprises two essential parts, namely the mixer unit and the linear-amplifier unit.

Fig. 10.—View of transmitter wing.

(5.5.1) Mixer Unit.

The function of the mixer unit is to translate a signal from the drive unit (centred upon 3·1 Mc/s) to the final frequency for subsequent amplification in the linear amplifier. Its design has been largely influenced by three main requirements in the transmitter specification, namely

(*a*) An output on any one of six spot frequencies in the range 4–27·5 Mc/s to be available in not more than 2 min.
(*b*) A very high performance in respect of spurious output frequencies.
(*c*) The crystal input frequencies to be within the range 3·4–7 Mc/s.

Items (*a*) and (*b*) have led to the use of six separate mixer units rather than a ganged one, which would give only a compromise performance over so wide a tuning range. Requirement (*c*) has necessitated the use of a harmonic generator in the oscillator input to the mixer, and (*b*) has demanded selectivity subsequent to the harmonic generator. Requirement (*b*) also suggested the use of a balanced mixer and dictated the degree of selectivity required after the mixer. Spurious frequencies are inevitable in any mixing system, and the system must be so designed that these fall outside the wanted band or are greatly reduced by selectivity and careful design of the mixer stage. The mixer unit is shown in block form in Fig. 11. Its spurious frequencies can be divided into two classes:

(i) Those removed (normally far removed) from the wanted frequency by a fixed amount.
(ii) Those which coincide with, or closely approach, the radiated signal at particular frequencies.

Spurious signals of class (i) are attenuated by selectivity between the mixer stage and the output, but those in class (ii) cannot be reduced by post-mixer selectivity. Consideration indicates that

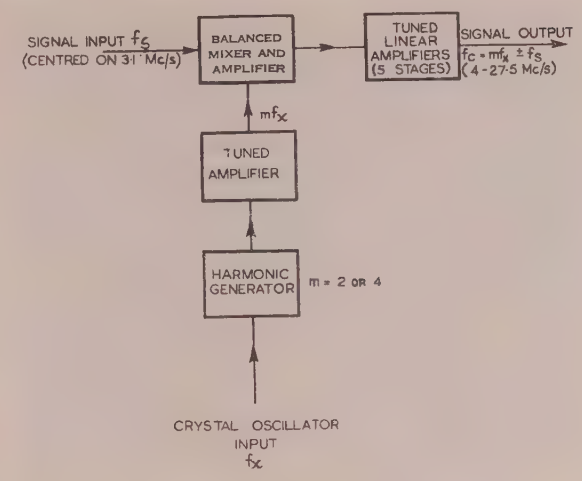


Fig. 11.—Block schematic of mixer and linear amplifier.

many spurious signals of this latter class are greatly reduced by the use of a balanced mixer. Certain other products which arise when the signal fed to the mixer contains multiples of the crystal frequency other than that desired are reduced to a very low level by including a selective stage between the harmonic generator and the mixer. The remaining spurious emissions are generally high-order terms, and are greatly reduced by approximating to square-law mixing and by maintaining a large ratio between the amplitudes of the crystal-harmonic signal and the 3·1 Mc/s input signal.

The use of six separate mixer units means that the size of each unit has to be limited. However, when components are mounted near to each other and to the chassis, appreciable changes of reactance may occur when large ambient-temperature changes take place. This would lead to loss of gain and to changes in the frequency response of the tuned stages; it has been largely nullified by connecting a capacitor of negative temperature coefficient across the anode inductor of each stage.

The 3·1 Mc/s signal is fed to the six mixer units in parallel, each unit being pretuned to one of the spot frequencies required, and the desired unit is selected by switching the h.t. supply and the low-impedance output circuit, the operation being part of the automatic frequency-change process. The crystal-oscillator input is fed to a harmonic generator at 100 mW and the selected harmonic passes to a balanced mixer. The output of the mixer is tuned to the lower sideband when the radiated frequency is below 10 Mc/s and to the upper sideband when it is above 10 Mc/s, and the output at the final frequency is fed to a balanced 100-ohm line.

The performance of the mixer units is summarized below.

Spurious outputs.—There are no spurious components of an amplitude greater than −56 dB relative to p.e.p. output, except for the second harmonic of the radiated frequency and the crystal harmonic frequency applied to the mixer, which are respectively at −40 and −50 dB relative to p.e.p. output. Since these products are far removed from the wanted frequency, their amplitudes are adequately reduced by subsequent selectivity in the linear amplifier.

Distortion.—Third-order intermodulation products are at least 46 dB below either of two equal test tones (1 100 and 1 775 c/s) at the normal working level, as shown in Fig. 12.

Gain stability.—The variation of output level over a period of six hours, with an ambient temperature variation of 15–50° C and a mains variation of ±1%, is less than 1 dB.

(5.5.2) Linear-Amplifier Unit.

The linear amplifier (Fig. 13) consists of five stages (Nos. 4–8). The output circuit of the 3-stage amplifier associated with the

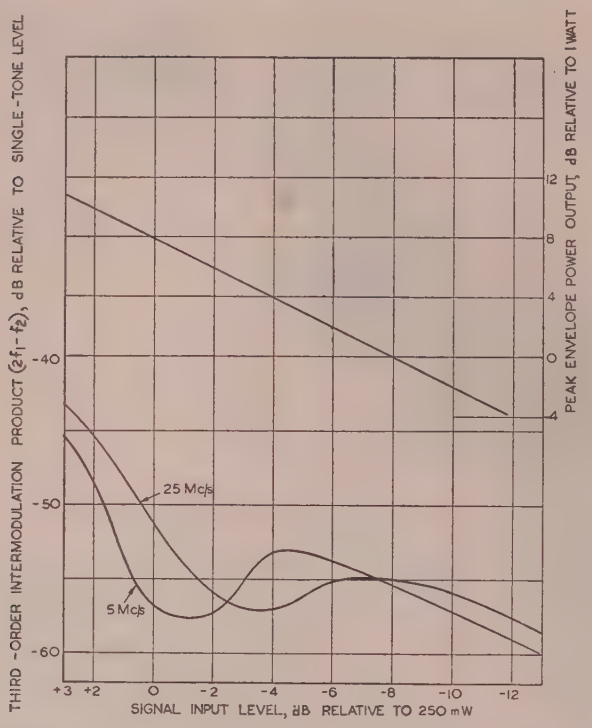


Fig. 12.—Third-order intermodulation product ($2f_1 - f_2$) from 3-stage amplifier following mixer.

with capacitance neutralization. It was apparent, however, that such circuits had little advantage, since valves of suitable construction were available for unbalanced earthed-grid operation. The unbalanced earthed-grid circuit has the advantage of not requiring neutralization in any form, while the number of valves, components and circuit adjustments is reduced to a minimum. Furthermore, the circuit as applied in the transmitters has fewer degrees of freedom to parasitic oscillations than a push-pull circuit of any type. The decision permitted the use of only one valve in each amplifying stage.

The characteristics of the valves used in the last three stages are given in Table 2.

Since the specification called for rapid adjustment of the transmitter to six predetermined frequencies by remote control, it was decided that the requirements could best be met by circuits built with continuously variable inductors and capacitors. The continuously variable inductor (Fig. 14) is wound with 0·75 in diameter high-conductivity copper tube on a Micalex former and has a spherical sliding contact held against the turns by a beryllium-copper spring. The turns are coated with Aquadag compound, the black surface of which ensures maximum heat radiation and lubricates the sliding contact. This type of inductor, evolved over the last 15 years, has proved extremely reliable.

The design of a continuously variable capacitor presented a much more difficult problem. Air-dielectric capacitors were ruled out by their bulk and high self-inductance, and a capacitor was therefore developed in which sulphur hexafluoride, a stable and inert gas, is used as a dielectric. This gas has an electric strength at 30 lb/in^2 equal to that of nitrogen at 200 lb/in^2. Coaxial electrodes (Fig. 15) are used to provide the maximum

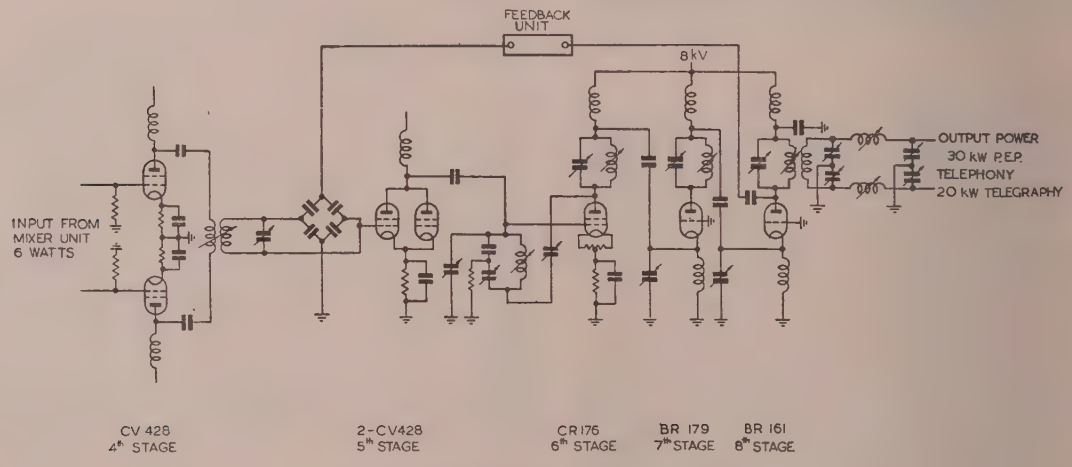


Fig. 13.—Simplified circuit of linear amplifier.

mixer unit drives stage 4, which consists of two CV428 valves in push-pull. The anode circuit of this stage is inductively coupled to two CV428 valves in parallel, which form the fifth stage. The circuit design of this stage is orthodox and does not need description here. The anode circuit of the fifth stage is coupled to the control grid of one CR176 tetrode through a capacitor.

The design of stages 6, 7 and 8, together with the main output circuit and harmonic filter is of interest. In view of the range of frequencies over which the transmitters are required to operate it was decided to use an unbalanced earthed-grid circuit in stages 7 and 8. Consideration was given to the use of circuits of the push-pull type with earthed grids or earthed cathodes, variation in capacitance with the minimum inductance. The vanes are contained in a Vitrosil cylinder protected by a shrunk-on polythene sleeve. The end caps are of Invar fixed to the Vitrosil by Araldite cement. The containers are permanently sealed, and when filled to a pressure of 35 lb/in^2 retain the gas pressure indefinitely. This form of capacitor is much smaller than any other type of equal rating, and its inductance is sufficiently low to permit its use as a harmonic by-pass element in the harmonic-rejector circuits of the transmitter. The variable inductors and capacitors described above are used in the last three stages.

The valve and circuit components of stages 6, 7 and 8 (Fig. 13) are contained in aluminium cabinets, the valves being mounted

Table 2

CHARACTERISTICS OF VALVES IN LAST THREE STAGES OF LINEAR AMPLIFIER

Detail		Stage		
		Sixth	Seventh	Eighth
Valve type		CR 176 tetrode	BR 179 triode	BR 161 triode
V_f, volts		5	6·6	9
I_f, amp		64	90	170
V_a (max), kV		8	8·5	12
P_a, kW		3·5	8–12*	15*–20
P_{g1} (max), watts		25	600	1 000
P_{g2} (max), watts		200	—	—
I_k (peak), amp		10	16	45
V_{g2} (peak), kV		1·5	—	—
μ		—	25	45
g_m, mA/volt		7	8·5	23
(Conditions of measurement of μ and g_m)	V_a (kV)	3	5	9
	V_{g2} (kV)	1·5	—	—
	I_a (amp)	1	1	1·2
f (max) at full rating, Mc/s		30	100	50
C_{gk}, $\mu\mu$F		40	40	57
C_{ag1}, $\mu\mu$F		0·4	30	36
C_{ak}, $\mu\mu$F		14	1	1·5

* Actual value depends on air flow.

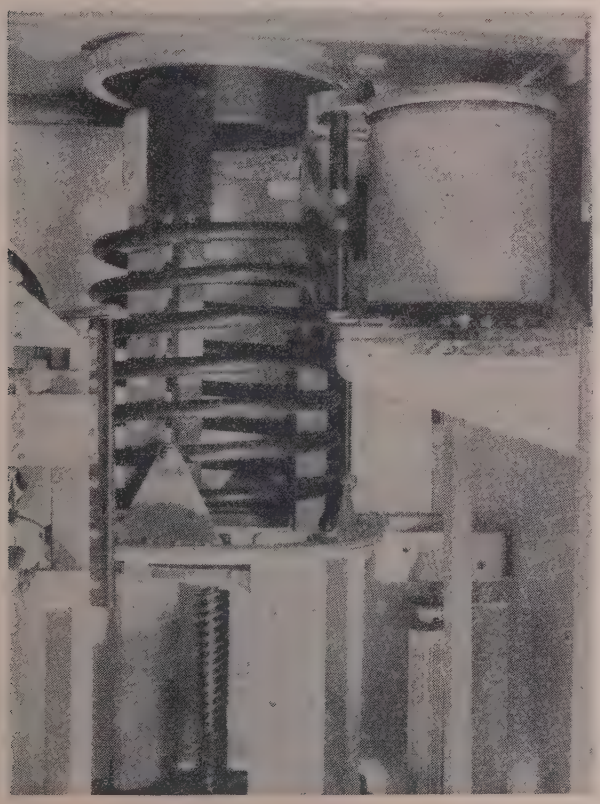


Fig. 14.—View of r.f. circuit showing variable inductor and capacitor.

anode downwards. The anodes are supported by a Micalex shelf, which acts as insulator and air baffle, the anode circuits being mounted directly below the valve. Above the anode shelf, at the level of the valve grid-seal, is a metal screen to which the grid is earthed through blocking capacitors. The anode is

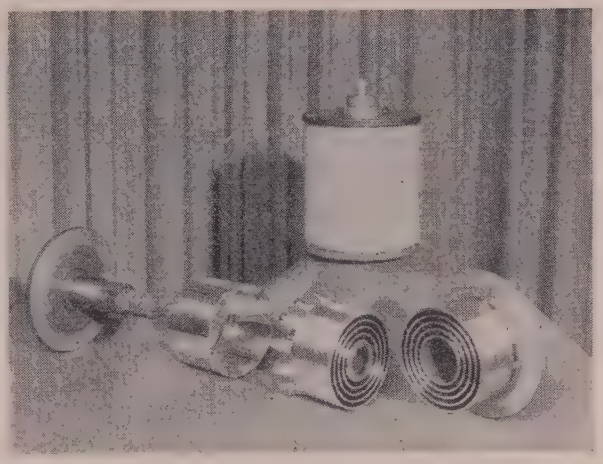


Fig. 15.—Variable capacitor.

thereby screened from the cathode, and a low-impedance path is provided for the heavy current circulating between grid and anode. The cathode circuits are mounted above the metal shelf and consist only of the cathode choke, wound with Pyrotenex cable. The compact circuit layout tends to keep stray r.f. currents from the frame of the transmitter and reduces radiation from the heavy-current circuits.

The output power of the penultimate amplifier is of the order of 2·8 kW; of this, 2·3 kW is transferred to the main output circuit by electronic coupling, and the balance drives the grid of the final stage under peak conditions. The output power of the final stage is 20 kW on f.s.k., or 30 kW p.e.p. on i.s.b., working.

Air cooling is used throughout the linear amplifier, air being drawn through a filter mounted at the top of the circuit screening boxes, through the valve anode fins and out through an aperture at the base of the box. The air stream thus cools the circuit components and ensures that cold air impinges on the filament and grid seals, thereby dispensing with the need for auxiliary cooling at this point. The total air flow through the valve fins is of the order of 2 500 ft^3/min at a velocity of about 80 ft/sec. The warm air from the base of the transmitter passes through a channel in the floor to the exhaust fan in its adjacent cubicle. After passing through the fan the air is either discharged outside the building or is used to heat the transmitter building. The doors of the circuit boxes are hermetically sealed; these boxes are mounted inside the main structural cubicles, the doors of which are also sealed and packed with sound-absorbing material, so that the noise 4 ft in front of the transmitters has been reduced to a very low level (about 60 phons).

The transmitter-output and feeder-coupling circuits were designed to function not only as impedance transformers but also as efficient harmonic filters. When operating at full power on any working frequency, the total harmonic power is generally less than 20 mW, and the highest value measured is 100 mW.

Negative feedback is applied from the anode circuit of the final stage and operates over the last four stages, the results showing that it is very effective in reducing non-linear distortion. Thus at maximum power, without feedback, the third-order intermodulation products are −28 dB relative to either tone level of the 2-tone test. With feedback these products are better than −36 dB relative to either tone at all levels up to peak envelope power.

The specification called for a transmitted signal/noise ratio of 60 dB. To ensure that this figure is achieved the cathodes of the valves in the sixth, seventh and eighth stages of the linear

amplifier are heated by d.c. power derived from 3-phase bridge-connected selenium rectifiers.

The high-voltage supply to these stages is provided at 8 kV by type AR63 grid-controlled mercury-arc rectifiers. The grid-control facility allows power to be switched on in steps of one-third, two-thirds and full voltage, and high-speed overload tripping with automatic resetting by Post Office type relays operates on the control grids. The circuit of the h.v. transformer primary is controlled by a 3-pole contactor with h.r.c.-fuse protection. The oil-filled power equipment is mounted in an individual fireproof cubicle adjacent to the transmitter.

A system of mechanical interlocks is incorporated to provide full protection for the operating staff. Limited access is allowed without switching off the filaments, but no points carrying dangerous voltages can be reached.

(5.6) Central Control Position

Operational control of the station is effected at the central control position. The function of the control officer is to ensure that the services required by the terminal stations in London are correctly and quickly provided. To this end he is equipped with remote-control and monitoring facilities, direct communication by order-wire to the London terminals, and communication by loud-hailer to staff patrolling the transmitter wings and by telephone to various strategic points. Remote-control facilities are provided for each transmitter, and the control unit also incorporates several alarms to warn the control officer when the corresponding items of monitoring equipment indicate unsatisfactory performance. Thus, from the information available at the c.c.p. the control officer knows at any time whether or not the station plant is functioning correctly, and if a fault develops he is made generally aware of its character and location.

Each transmitter has a separate control unit, and the 30 units (two spares are provided) are accommodated in five sloping-fronted metal cabinets mounted on a central control desk. This desk, shown in Fig. 16, is U-shaped and some 14 ft wide, with

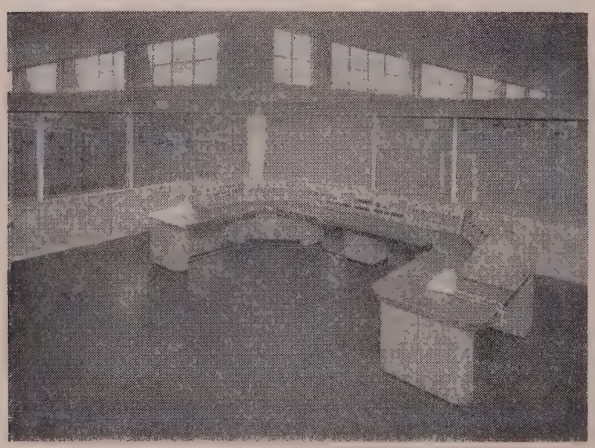


Fig. 16.—View of central control position.

limbs 8 ft long. The control units are identical and interchangeable, only their cover plates differing according to the transmitter. Any unit can be quickly removed for inspection or replacement with a spare by pulling off the cover plate and lifting the unit from the bottom to disengage the connecting jacks. The centre of the control desk houses two volume indicators and a speaker-circuit control panel.

Pressing the "XMTR ON" button at the c.c.p. initiates the starting sequence at the transmitter, and its completion is signalled by illumination of the transmitter-number lamp at the top of the control unit. Pressing the "XMTR OFF" button shuts the transmitter down completely. The frequency-selector switch can be set to select any one of the six preset frequencies of the transmitter. This switches off the h.t. supply at the transmitter and starts the frequency-changing mechanism, which is controlled by cam-operated switches fitted to the transmitter-tuning motors; these switches automatically restore the h.t. supply to the transmitter when the tuning operation is complete. A switch at the c.c.p. transfers control from the c.c.p. to the transmitter itself for maintenance work. At the transmitter another switch can be used to set the frequency-change circuits to manual control; operation of this switch returns an indication to the c.c.p. A switch is available for changing from "Traffic" to "Local Control" for testing.

The transmitter control units on the c.c.p. include indicators which show whether the pilot-carrier level is correct, and whether the reflectometer indication of aerial impedance is normal. In addition, lamps show for each sideband whether the automatic monitor checks of the outgoing telephone or telegraph signals are satisfactory.

(5.7) Aerial Switching Equipment

Twin coaxial feeders having a characteristic impedance of 194 ohms balanced to earth carry the power from the transmitters to one or the other of two aerial switches located in separate aerial switch rooms. From the switches the transmission lines, still in the form of twin coaxial feeders, pass to the rear parapets of the two transmitter wings that flank the central hall, and thence via exponential matching lines to the open-wire transmission lines. Fourteen transmitters and up to 36 aerials can be connected to each of the two switches, which are interconnected by junction feeders so that arrangements can easily be made for any aerial to be energized by any transmitter.

Each of the two aerial switches comprises an array of 14 "switch decks", one of which is shown in Fig. 17. The switch decks are mounted horizontally one above the other (Fig. 18); each deck is associated with a particular transmitter and enables it to be connected automatically to one of six preselected aerials. In fact, since only 36 aerials will be available to the 14 transmitters joined to one switch, the number of preselected aerial positions provided on most of the switch decks is about three. The switch deck comprises horizontal twin-coaxial feeders joined to the transmitter, with side outlets for the aerial connections. The 36 aerial twin-feeders are terminated in telescopic sections with elbows at their ends and hang down vertically on either side of the stack of switch decks. On any particular switch deck the side outlets are fitted in the correct positions to mate with the elbow connectors of the feeders to the associated aerials. The inner conductors of the horizontal feeders in the switch deck are interrupted by internal switches at each of the aerial side outlets; these switches either divert the feeders to the outlets or switch them on to the next pair of outlets, and they are operated under remote control by a motor at the end of the deck. The selection of the aerial is automatically controlled by the 6-position frequency-selector switches at the central control position, and is completed in less than one minute.

Additional facilities are provided to enable a transmitter to feed a switch deck not normally associated with it, and this, with the limited amount of interconnection previously mentioned, gives complete flexibility. The switch-deck components are interchangeable, and spare decks are provided to reduce the interruption time when setting up new switching arrangements.

The feeder tubing is of high-conductivity copper; the inside diameter of the outer and the outside diameter of the inner

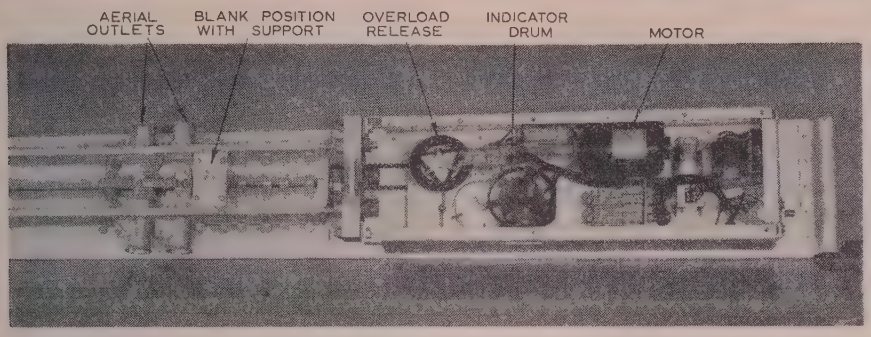


Fig. 17.—Aerial switch deck (cover removed).

conductors are 1·5 and 0·3 in respectively, the inner conductor being supported by 3-leg ceramic insulators soldered to it at 1 ft intervals. Expansion joints are provided in the core only, to allow for differential expansion between the core and sheath, the latter being free to expand as a whole in its supports. The overall attenuation of the tubular-feeder connections is substantially proportional to length, the introduction of the aerial switches and the several right-angle bends having little effect on the loss. The loss is approximately 0·063 dB per 100 ft at 10 Mc/s and is closely proportional to the square root of the frequency. The modulus of the reflection coefficient at the transmitter end of a typical tubular feeder, when terminated at the far end in 194 ohms, is shown plotted against frequency in Fig. 19(*a*).

(6) EXTERNAL PLANT

As has been indicated, the site is intended to accommodate more than 70 directional aerials, most of which will be rhombics. The layout is, so far as possible, planned with the aerials spaced around the perimeter of two roughly circular areas north and south of the station buildings, the object being to provide each aerial with an unobstructed foreground in the direction of maximum radiation. The initial provision totals 58 aerials, laid out as shown in Fig. 2, and comprises 18 small rhombics for the easier circuits, 28 large rhombics for the more difficult

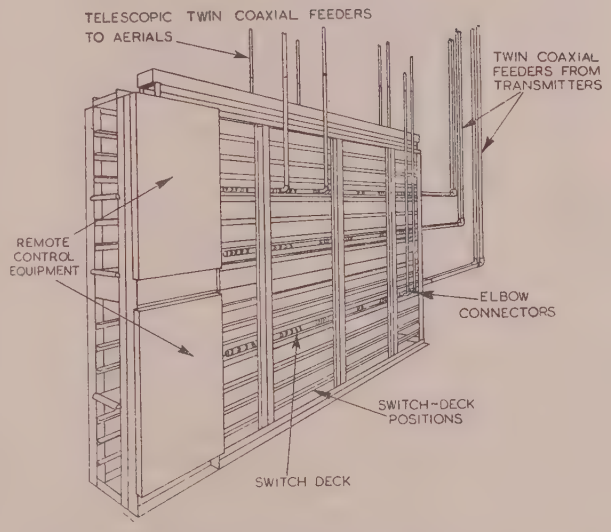


Fig. 18.—Aerial switch showing three switch decks in position, two being connected to two aerials.

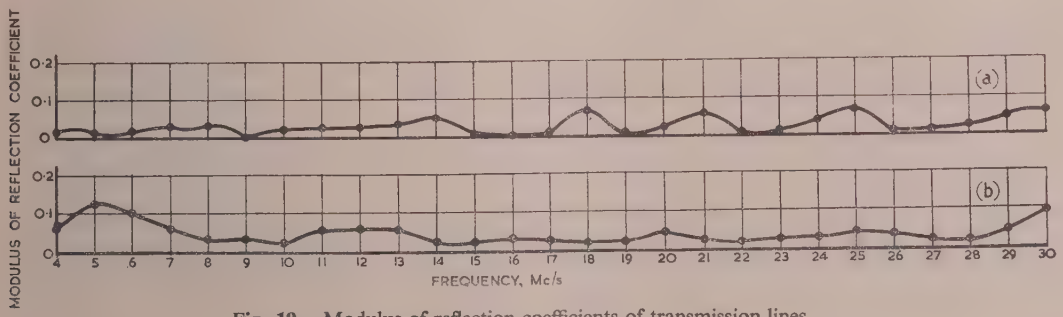


Fig. 19.—Modulus of reflection coefficients of transmission lines.

(*a*) Typical twin coaxial feeder with far end terminated in 194 ohms.
(*b*) Exponential matching line with far end terminated in 565 ohms.

circuits, a set of four Kooman's arrays for New Zealand, to cater for the unusual requirements of this circuit to the antipodes, and eight quadrant dipoles for emergency and omnidirectional use.

The transmission lines feeding the aerials leave the station buildings in two main routes, one from each of the two aerial switching rooms. Some seventy 80–90 ft spliced wooden poles, fifty-five 150 ft and three 325 ft steel lattice masts, and nine hundred 24 ft wooden poles have been erected on the site. Approximately 30 miles of 2-wire transmission line have been run.

Before construction commenced the site was surveyed and triangulation points forming a complete network over the site were permanently marked by concrete obelisks. By means of solar observations the true geographical north was established.

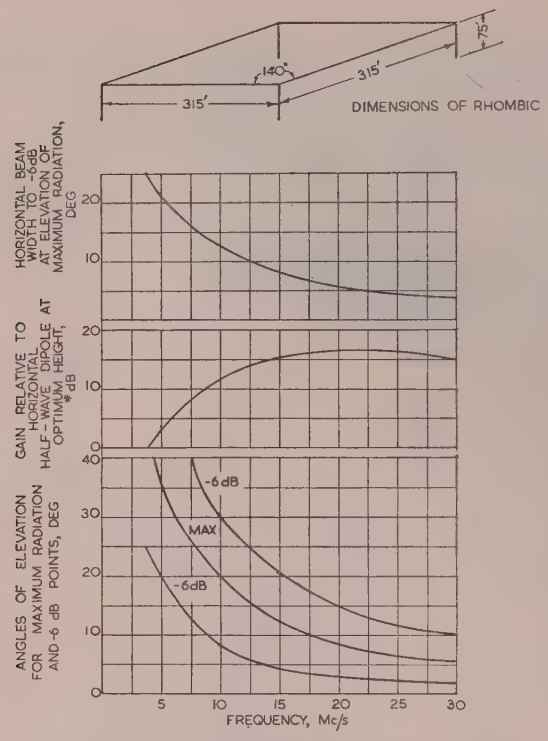


Fig. 20.—Characteristics of small rhombic aerials.

* This gain is 6 dB less than the gain referred to a dipole in free space.

Thus the network forms a permanent basis for setting out aerial systems.

All the rhombics are of the triple-wire type, and both small and large ones are provided. The small rhombics, whose physical and radiation characteristics are shown in Fig. 20, are used only for those circuits where gain at low angles of elevation and at the lower end of the h.f. band is not of prime importance. On the other hand, for circuits where high gain is required, particularly at the lower frequencies and at low angles, pairs of large concentric rhombics are normally provided, mounted on four 150 ft steel masts, one of a pair being 75 ft, and the other 150 ft, above the ground. The side-lengths of the two large rhombics in a pair are nearly the same, but the side-angles differ somewhat, the lower and the higher aerials being designed for use in the upper and lower portions of the h.f. band respectively. Each aerial of the pair covers a frequency range of about 2·5 : 1, and between them they cover virtually all the frequencies likely to be used on any one service. The large rhombic aerials so far needed fall into three categories:

(*a*) Those for the North Atlantic route, which are the largest, because the difficult propagation conditions and the importance of the circuits make high gain particularly desirable.

(*b*) Those for Australia and Singapore, which are exceptional in two respects: only one design of aerial is needed to cover the restricted range of frequencies that can be used on these services, and there is a requirement for reversing the direction of shoot to provide for long- as well as short-path transmission. However, for each service two aerials are mounted on the same set of masts, at heights of 150 and 112·5 ft, for use when two transmitters are operating.

(*c*) Those for certain other services.

Certain basic performance data on these three designs are given in Fig. 21. The top curves show how the horizontal beam-width of the major lobe varies with frequency; those in the

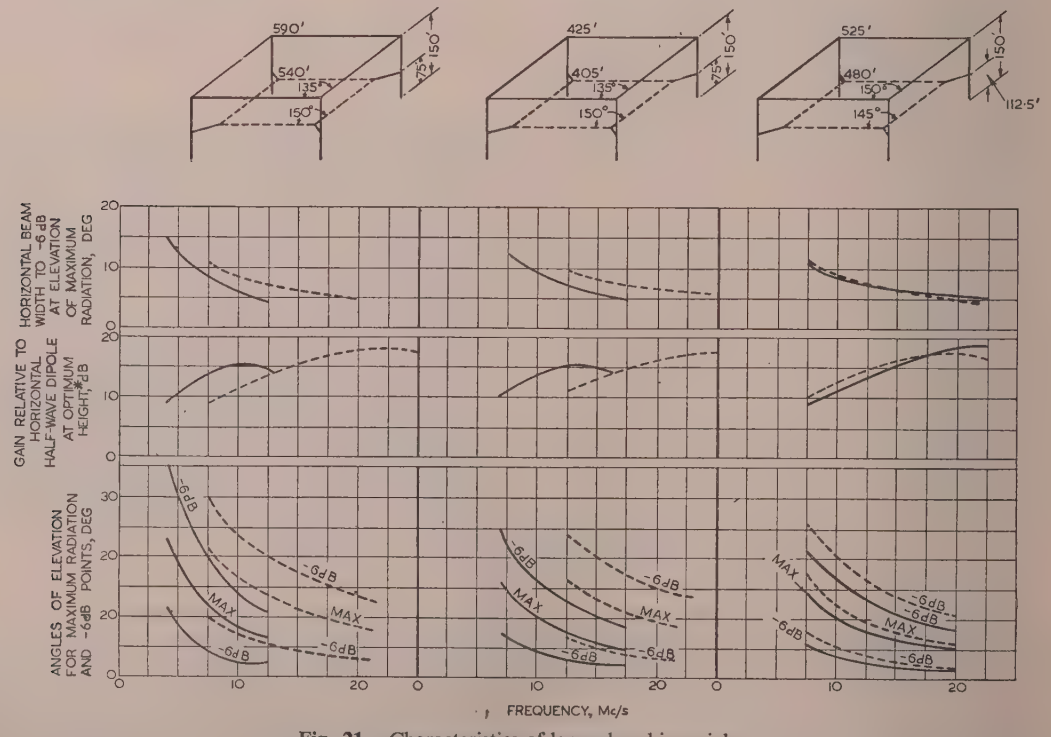


Fig. 21.—Characteristics of large rhombic aerials.

* This gain is 6 dB less than the gain referred to a dipole in free space.

centre give the gain/frequency characteristics; the bottom curves show how the directions of maximum radiation and of the 3 dB points vary in the vertical plane as the frequency changes.

Experimental observations of the direction of arrival in the United Kingdom of transmissions from New Zealand have shown that wide deviations from the great-circle path occur, and suggest that a coverage of $\pm 30°$ on a mean bearing of 30° east of north is desirable. For this reason, Kooman's-type aerials are provided, each consisting of one vertical stack of full-wave horizontal dipoles with reflectors, the mean height of each aerial being arranged to embrace the desired vertical angles of propagation. Four aerials are provided to cover the range of frequencies that will be needed during the sunspot cycle, and three 325 ft stayed masts are used to support them, two arrays being suspended, one above the other, between adjacent masts. The reflector and radiator curtains are identical and their connections can be interchanged by remotely controlled switches to provide for long- or short-path propagation.

The omnidirectional aerials are essentially similar to those described by Wells,[9] but are specially reactance-compensated[10] to improve the standing-wave ratio.

Iron-wire dissipative lines used for terminating the rhombics are run along the major axes of the latter 15 ft above ground, the far ends being left open to facilitate insulation tests. The attenuation of the lines is 12 dB at 4 Mc/s.

From the aerials, 2-wire transmission lines of 565 ohms characteristic impedance converge into the two main routes, which follow existing hedgerows as far as possible to minimize disturbance to agriculture. Near the building the open-wire transmission lines run via the exponential matching lines to the 194-ohm twin coaxial feeders, and thence to the aerial switches. The coaxial feeders on the roof are connected at the parapet to the exponential lines, which extend for a distance of 135 ft to the main-line-terminating gantry at which the exponential and 2-wire lines are connected. Each exponential matching line uses four wires, the horizontally adjacent wires being in parallel; the lines are uniformly tapered in elevation, and, with only two changes of taper in plan, a substantially exponential impedance taper is obtained. The line is simple and easy to construct, and its performance, when perfectly terminated, is given in Fig. 19(*b*).

The transmission lines on the main routes are built on wooden gantries. Low-capacitance post-type insulators, through which the wires run freely, are used to support the lines at most poles. At about every sixth pole, however, the wires are mechanically held in low-capacitance tension insulators, The result is a line of excellent uniformity.

Each open-wire line consists of two 400 lb/mile copper conductors, one 9 in above the other, supported at a mean height of 16 ft above ground. On the main routes the pairs are arranged side by side, with a separation of 2 ft. The results of measurements of attenuation and crosstalk on such lines are given in Table 3.

Care has been taken in the mechanical design and construction of the aerials and lines to obtain good electrical performance and low maintenance costs. The rhombics are constructed of light cadmium-copper wire (70 lb/mile), thus enabling the sags to be kept small and light masts to be used. The additional tension due to wind and ice loading is limited by making off one of the supporting ropes to a weight which is lifted from the ground when the maximum safe loading is reached. The wires are anchored in low-capacitance insulators to avoid the localized capacitance of binding wires, which constitute appreciable electrical discontinuities at the apices of a rhombic. The design of the throats of the multi-wire rhombics has also received attention, since at these points the taper of the wire spacing critically affects the uniformity of impedance.

Table 3

ATTENUATION AND CROSSTALK MEASUREMENTS ON THE OPEN-WIRE AERIAL FEEDERS

Frequency	Attenuation of 1 500 yd of open-wire line	Crosstalk attenuation	
		At station end between two adjacent 700 yd lines and rhombic aerials on common masts	At aerial end between two adjacent 1 500 yd lines, near end terminated in Z_0
Mc/s	dB	dB	dB
4	1·1	35	43
10	1·5	31	36
20	2·2	34	27
30	2·8	30	22

The 150 ft masts supporting the large rhombics are of light welded-steel construction, galvanized overall. Although they are triangular in section the posts are of standard angle braced with solid rod. Each mast weighs only 1 500 lb and can safely withstand loads of one ton applied horizontally at two stay points.

(7) POWER-SUPPLY ARRANGEMENTS

The power-supply arrangements for the new station are outlined in Fig. 22. From the old station, power drawn from the 11 kV mains or from an emergency supply is conveyed to the new station by two well-separated underground feeders, each of which is capable of carrying the full load of about 1 400 kVA. The feeders supply two interconnected substations behind the transmitter building, and by a local diversion en route one of the feeders also supplies certain incidental loads at the testing and transmitter-development block.

Owing to a limitation on the maximum permissible fault-clearance time imposed by the supply authority, a balanced-current feeder-protection system has been adopted in which the feeder system is divided into overlapping zones, any one of which is automatically isolated in the event of a fault occurring in it, without affecting the supply to the others.

The load is shared between the two substations, each of which contains three 500 kVA transformers, two working and one spare, for supplying the medium-voltage distribution system. The transmitters are supplied in groups from two m.v. switchboards in the transmitter wings, and since the grid-controlled mercury-arc rectifiers in the transmitters will clear overloads faster than any circuit-breaker, the m.v. feeders have to be protected against catastrophic failure only. This is achieved by the use of h.r.c. fuses both in the grouped and in the individual supplies to the transmitters. Lighting and miscellaneous small-power requirements, and a regulated supply for the equipment in the rack-apparatus room, are met from an auxiliary switchboard fed from the m.v. switchboards.

The emergency supply is provided by Diesel-engine generators in the old station. A 1 400 kVA set was already available there, and in the interests of building and staff economy it was decided to provide two more 900 kVA sets there, rather than to establish a separate emergency supply at the new station.

All oil-filled equipment is installed in fireproof rooms provided with carbon-dioxide protection.

(8) PERFORMANCE

The detailed performances of the component parts of the transmitters have been described already in earlier Sections of the paper. Experience in operating the station since it was brought into limited service in October, 1954, has clearly con-

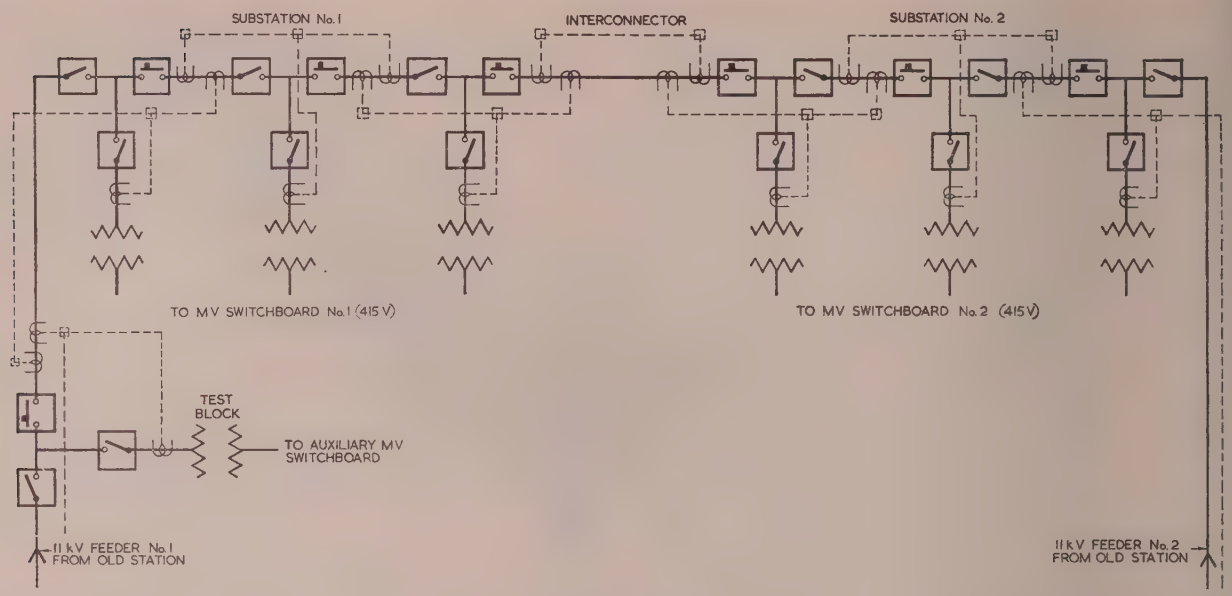


Fig. 22.—Power supplies at the new station.

——— Power cables.
— — — Protection system.

firmed the advantages of providing a common type of high-power linear amplifier for use with either telephone or telegraph drive units. This flexibility has proved valuable, as also has that provided by the aerial-switching arrangements.

The use of air cooling for the high-power stages has obvious operating advantages, and air noise has been kept to a very low level. The remote-control and monitoring facilities from the central control position have worked well and should lead to substantial savings in operating costs.

(9) ACKNOWLEDGMENTS

The authors would like to thank the Engineer-in-Chief of the Post Office and the General Manager of Marconi's Wireless Telegraph Company, Ltd., for permission to publish the paper. The provision of the new station has been made possible only by the loyal assistance of many colleagues, and to them the authors express their appreciation. Thanks are also due to Mr. H. Stanesby and Mr. W. J. Morcom for their advice on the preparation of the paper. Finally, the authors acknowledge the very effective co-operation of the Ministry of Works in the provision of the buildings.

(10) BIBLIOGRAPHY

(1) West, W., Cook, A., Hall, L. L., and Sturgess, H. E.: "The Radio Transmitting Station at Criggion," *Journal I.E.E.*, 1947, **94**, Part IIIA, p. 269.

(2) Shaughnessey, E. H.: "The Rugby Radio Station of the British Post Office," *ibid.*, 1926, **64**, p. 683.

(3) Hansford, R. V., and Faulkner, H.: "Some Notes on Design Details of a High-Power Radio-Telegraphic Transmitter using Thermionic Valves," *ibid.*, 1927, **65**, p. 297.

(4) Oswald, A. A., and Deloraine, E. M.: "Transatlantic Telephony," *Electrician*, 4th June and 25th June, 1926.

(5) Bray, W. J., Lillicrap, H. G., and Lowry, W. R. H.: "The Design of Transmitter Drives and Receivers for Single-Sideband Systems," *Journal I.E.E.*, 1947, **94**, Part IIIA, p. 298.

(6) Law, H. B.: "Standard Frequency Transmission Equipment at Rugby Radio Station," *Proceedings I.E.E.*, Paper No. 1762 R, October, 1954 (**102** B, p. 166).

(7) Creighton, J. L., Law, H. B., and Turner, R. J.: "Crystal Oscillators and their Application to Radio Transmitter Control," *Journal I.E.E.*, 1947, **94**, Part IIIA, p. 331.

(8) Booth, C. F.: "The Evolution of Frequency Control," *Proceedings I.E.E.*, 1951, **98**, Part III, p. 1.

(9) Wells, N.: "The Quadrant Aerial: An Omnidirectional Wide-Band Horizontal Aerial for Short Waves," *Journal I.E.E.*, 1944, **91**, Part III, p. 182.

(10) Zinke, O.: "The Principles of Broadband Aerials for Metre and Decimetre Waves," *Funk und Ton*, 1950, **9**, p. 437.

(11) *The Builder*, 23rd September, 1955, p. 518.

DISCUSSION BEFORE THE INSTITUTION, 3RD NOVEMBER, 1955

Mr. H. Stanesby: The paper describes what is, I think, the biggest high-frequency radio station ever to be built as a single project, and points the way to the future in many respects.

The outstanding features of the new station are the cleanness of the layout, the high degree of standardization and flexibility in the plant, and the saving in man-power and ineffective transmitter time resulting from the use of remote transmitter control combined with automatic frequency changing and aerial switching. In these days of full employment and high labour costs, productivity is of paramount importance, and clearly the designers have had this very much in mind.

The authors refer to an important technical requirement—which was also discussed in the paper on the design of the original Rugby station*—namely frequency stability. The

* Shaughnessy, E. H.: "The Rugby Radio Station of the British Post Office," *Journal I.E.E.*, 1926, **64**, p. 683.

present international tolerance for high-power point-to-point h.f. transmitters is ± 30 parts in 10^6, which allows the band occupied by a transmission on 20 Mc/s to vary in frequency by 1 200 c/s. This is far more than that needed by any one of many different types of radiotelegraph transmission in use to-day, if the transmitter is well designed and the frequency is held constant. If others were to follow the lead given in the paper, by increasing the frequency stability and reducing the spurious emission of the transmitters, the frequency spectrum could be used much more efficiently.

A great deal of interference with television reception is being caused by harmonics from h.f. transmitters; it is therefore very important that such radiation should be reduced to very low levels. Harmonic radiation from the aerials and feeders can be reduced by using filters in the transmitter output circuits, but the transmitters themselves may radiate if the screening is inadequate. The performance of the new Rugby station is very satisfactory in this respect. Preliminary measurements made at points one mile from the station buildings show that in only one case did the harmonic radiation in television Band I reach $10\,\mu$V/m. In all other cases it was $3\,\mu$V/m or less. This is to be compared with the field strength of $100\,\mu$V/m which is regarded as the minimum for satisfactory television reception. Such low harmonic radiation makes it very unlikely that the new station will interfere with television reception.

Mr. H. Faulkner: It will be remembered that about 30 years ago Franklin was very busy designing the first beam transmitters, which undoubtedly revolutionized long-distance communication.

Work on much higher ranges of frequency followed, and it became necessary for the C.C.I.R. to invent a new nomenclature for frequencies, in which the band of 3–30 Mc/s is known as Band No. 7. Why have the authors not used this nomenclature in the paper?

Thirty years ago, oil cooling was used for the transmitters. I remember the first small Post Office transmitter erected in Handley Cross farm on the Rugby site; oil was used to cool the valves, and proved a messy job and not at all pleasant. Fortunately, engineers began to think that Ohm's law applied equally well to water in pipes as to other things, and that the losses in water columns would not be very large, and so we got water cooling. We have now gone a stage further and have air cooling. I believe that the designers have made a good step forward in choosing air cooling for this purpose, and the use of some of the air to warm the building is also of some importance.

The method of frequency control adopted at this station and the segregation of the low-power circuits from the high-power amplifiers is of interest. The fact that it is possible to keep the transmitted frequency well within the C.C.I.R. standards by means of a simple crystal panel which does not require thermostatic control should encourage organizations throughout the world to come within these standards, to the great advantage of all.

The Post Office have long been in the van as regards frequency stability, and even the first long-wave transmitter installed nearly 30 years ago was provided with a tuning-fork control—the tuning-fork having a very low temperature coefficient—and at that time it was claimed to be the transmitter with the most constant frequency in the world.

The application of negative feedback to short-wave transmitters is of great importance in providing stability and linearity, and it is remarkable what advances in broad-band amplifiers have been obtained by this simple device.

The central control position provided in the station is a practical development which effects considerable savings of staff and consequent reduction in operating costs. The possibility of changing frequency and aerial connections by pushbutton control from the central position must have needed a great amount of engineering development and is a big step forward.

The project is an example of co-operation between the user, the Post Office and the manufacturer. None could have done the job so well alone, and this close co-operation is most valuable. It is not unusual in the Post Office, and it applies in the line telephone and telegraph fields as well. It is something which we should carefully nurture and preserve, since it is most valuable from the point of view of the country as a whole.

Mr. J. A. Smale: The grouping of transmitters or transmitter-amplifiers in unattended buildings has been adopted at several British stations, with considerable advantage in the cost of feeders and their losses. Table 3 indicates an average loss of about 1 dB per transmitter; a further third of the power is dissipated by the rhombic aerials, giving an overall efficiency of about 50%. With 20 transmitters working at 20 kW for, say, 12 hours a day, some 800 000 kWh are dissipated each year. As the cost of generating h.f. power is about 2s. 6d. per kilowatt-hour, this represents an annual loss of about £100 000. A partial saving would help to pay for more efficient aerials and other modifications in the station design. However, in view of the total running costs, it may not be too much to pay for relative simplicity and flexibility.

The authors should say why air cooling was chosen; I agree that it has obvious operational advantages for high-power stations, but it also has disadvantages, particularly where high powers are concerned and large losses must be removed by the cooling system. Prominent among these is the formation of scale on the valve anodes and the consequent difficulty of controlling the hardness of the water.

In Section 3.2 the authors say that "the highest gain compatible with reasonable cost" is desirable in an aerial. A broadside array gives some 6 dB of additional gain and can be used with one-quarter the power; combined with other facts given in the paper, I realize why this choice was made, and agree that, in the circumstances, it was wise.

I am pleased to see that f.m. as well as a.m. continuous wave is classified as A2: there has been a tendency to call the former F2, whereas at the receiver it is indistinguishable from the latter.

In Section 5.2.2 it is stated that "For single-channel f.s.k. working the incoming d.c. keying signals can be connected to a signal-inverter stage. This is used to correct the frequency inversion that takes place in a subsequent mixer unit. . . ." Is this done automatically? If not, human error can produce a chaotic state of affairs. Could not a transmitter be designed in which this was unnecessary? This reversal of sideband signals when the h.f. part of the transmitter is working above or below 10 Mc/s has been a nuisance.

In Section 5.4 the authors say, "The telephone monitors compare the envelopes of the land-line and transmitter-output signals in terms of the charges developed on two *RC* combinations. . . ." This is satisfactory for telephone channels, but what happens when the telephone channel is carrying a number of telegraph channels, one of which may be in faulty adjustment?

In Section 5.5.2 it is said that the continuously variable inductor has been evolved over the last 15 years. It would be truer to say that it has been a nuisance over this period, especially at the higher frequencies, and that it is only very recently that a solution has been found.

In Section 5.5.2 reference is made to the grid-control facility in mercury-arc rectifiers which allows power to be switched on in three steps. I consider this facility unnecessary; what is more important is whether it is possible to leave the transmitter operating in one of these starting-up positions, because if so

there is a danger of serious mains hum, for grid control on the mercury-arc rectifier produces an extremely bad waveform.

Colonel D. McMillan: As a user of this station, I feel that our experience to date indicates plainly that all we had hoped for in the design is being achieved. We are so delighted with it that the future development and expansion we have in mind will doubtless follow the same plan, i.e. the standardization of the equipment and a cruciform layout of building with a central control point, with the advantages which that gives—notably in the saving of specialized staff.

The new Rugby station, in conjunction with the new receiving station at Bearley, has, at least for the present, brought us to the point where shortage of major radio equipment is no longer a deterrent to the commencement of new services. The real deterrent now is the difficulty with frequencies. Attention has been drawn to the contribution which this station makes towards the most advantageous use of those frequencies which we have, and this is probably the only way in which we can face the difficulties of the future.

I was a little worried to notice in Section 3.1 a remark which might suggest that the exploitation of this new station for multi-destination services may tend to waste some of the advantages of frequency stability. I agree with the authors that such services put a premium on high-grade high-power transmission, so that relatively inefficient receiving equipment can be tolerated. We are well aware of the undesirable effects in a congested ether of that type of engineering, and we shall try to persuade some of our customers to adopt point-to-point transmission for the "main haul" of such services and to resort to the broadcast part only for the local distribution at the far end. How successful we shall be remains to be seen, but it is not a point which has been missed.

Another important characteristic of the new station from the operational aspect is the automatic monitoring equipment. A station such as this represents a considerable investment, and attention has already been drawn to the power wasted in heating the surrounding atmosphere, owing to the nature of transmitters; nevertheless avoidable waste must be eliminated, and the automatic monitoring equipment—I think for the first time in our experience—provides the opportunity of ensuring maximum transmitter efficiency at all times.

The station also contributes to the objective of using the investment to the best possible advantage by the arrangements for the rapid switching of the transmitters and aerials. The designers have made a courageous attack on the very difficult problem of aerial switching, and from our small experience so far we are all most hopeful that it will prove finally to have solved that problem.

Mr. A. N. Thomas: For long-distance services it is our experience in the B.B.C. that the advantage of the wide frequency coverage of the rhombic aerial is largely lost because the vertical angle of maximum radiation varies with frequency. The paper indicates that the average gain of the rhombics used approximately equals the forward gain of a stacked array of dipoles comprising 16 elements. The stacked array, however, has smaller side-lobe radiation and its direction of maximum gain can be slewed through a limited arc. However, one would assume that the transmitting station design was greatly assisted by the relatively complex receiving installation which can be made available, which reduces appreciably the minimum power required for effective communication over difficult circuits. Could the authors summarize the effectiveness of the maintained service to Singapore at times around noon G.M.T.? This is sunset in Singapore, and particularly difficult conditions are experienced with the reception of broadcast transmissions from this country. To maintain a service during this period (or to resume it as soon as possible) the B.B.C. has used a transmitter power of 100 kW working into a stacked array of 32 dipole elements plus reflector. Does the authors' choice of a stacked dipole array for the New Zealand service also reflect their faith in such arrays for difficult long-distance service?

For the aerial switching system a combination of cable-type feeders and open-wire lines has advantages. Cables are obviously a necessity for the type of feeder switching adopted, but they are less easy to maintain; what steps have been taken to put a static earth on the open wire feeder lines to keep lightning surges, etc., out of the cables? The B.B.C. would normally have $\lambda/4$ short-circuited stubs available to provide an earth connection and a degree of filtering, but this is clearly unacceptable where a number of frequencies are being carried on the same transmission line.

The use of a common drive-unit conforms with B.B.C. practice on multi-transmitter stations, although the drive we use is inherently simpler, but there are several small points on which I should like further information:

(*a*) Is thermostatic control of drive-room temperature necessary?
(*b*) What heat dissipation is considered satisfactory within the standard drive-room cabinet?
(*c*) Is there any information on the operating temperature of the valves?
(*d*) Does the provision of the frequency-measuring set mean that the station staff can observe with sufficient accuracy crystals of long-term stability better than 15 parts in 10^6, and does it mean that the frequency is not monitored elsewhere (e.g. at Banbury)?
(*e*) Is any information available on the fault record of the drive units since the station went into service?

The transmitters are specified for rapid adjustment to six predetermined frequencies by remote control, clearly achieved by preset adjustments; it would surely be difficult to set up the circuits to substitute one transmitter for another in an emergency unless the desired emergency frequency was included in the preset adjustments.

I note that crosstalk attenuation between neighbouring systems is better than 30 dB, which is presumably considered adequate because of the very good filter characteristic on the output of the transmitters. However, I should like some information on the measured level of spurious frequency radiation with neighbouring pairs energized. Under similar circumstances it has been our experience that spurious frequencies of the order of $2f_1-f_2$ have been the most troublesome, and 30 dB cross-talk attenuation between adjacent lines would not invariably be adequate. Are *ad hoc* filtering arrangements made and is this made more difficult by the use of open-wire feeder lines with pairs in the vertical plane? It is the absolute level of the spurious frequency which is objectionable, and relative levels of crosstalk attenuation cannot be directly related.

Mr. W. J. Morcom: The requirement of remote control in a transmitter which had to be tunable to any frequency in the band of 4–27·5 Mc/s, which had to match feeder impedances of 200 and 600 ohms with standing-wave ratios of 2 : 1, and yet preserve a high degree of linearity and gain constancy without an abnormal number of tuning variables, necessitated the adoption of techniques not previously employed.

The variable gas condensers were developed specifically for this project, and each of the three final high-power stages uses a variable gas condenser ganged to a variable inductor, so that only one spindle need be rotated to tune each of the anode circuits over the complete frequency range. By similar arrangements on lower-power stages, single-spindle control of tuning over the whole frequency range is also effected.

In order that servicing of the tuning motors and their associated preselector mechanism can be undertaken easily, the complete motor heads are mounted on detachable panels complete with plug-in electrical connections and dog clutches.

hese can be unscrewed from the front of the transmitter panels, ven, if necessary while the transmitters are still in operation, and an be taken away for servicing. Thus, in an emergency, service an be maintained even if there are faults on the automatic tuning ircuits.

Mr. A. Cook: There is some difficulty in determining the ptimum power to be provided at the transmitter, because his depends, among other things, on the performance of the eceiving equipment, the noise at the remote receiving station, he propagation path, the transmitting-aerial gain and the haracteristics of the method of signalling. There is thus a arge number of variables in the problem.

If we cannot arrive at a clear-cut calculation of the power equired, we can at least calculate the relative costs of generating he power and of providing aerial gain at the transmitting station. This has been done for several types of modulation, and Fig. A

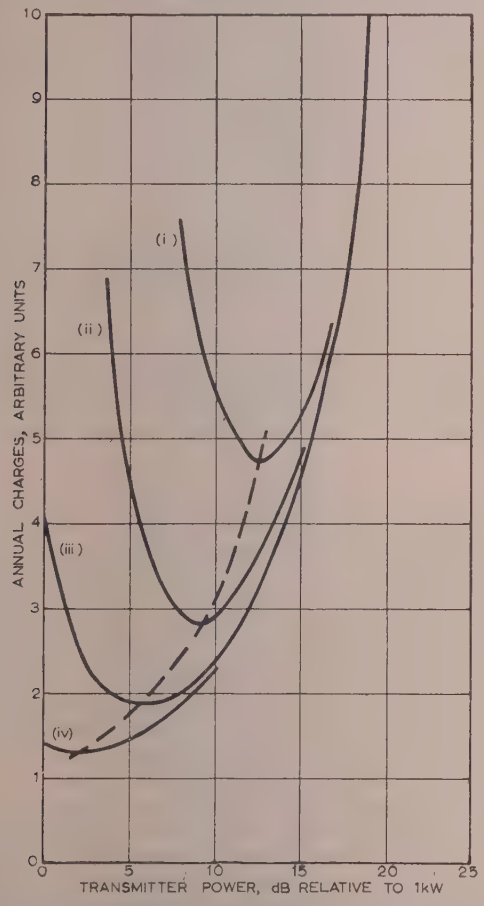


Fig. A.—Relationship between transmitter power and annual charges, or various effective radiated powers relative to 1 kW in a $\lambda/2$ dipole bove a conducting earth.

Telegraph service F_1: 7000 hours per annum:
(i) 30 dB e.r.p.
(ii) 25 dB e.r.p.
(iii) 20 dB e.r.p.
(iv) 15 dB e.r.p.

s one such calculation relating to a frequency-shift telegraph ervice operated for 7000 hours per annum. Costs in other rganizations may be slightly different, but I hope that publica- ion of this Figure will encourage others to produce data from heir own organizations, so that we may compare results. It will e noted that the minimum of each curve is fairly sharp, and I would stress that the immediate object of the calculations is to show whether we have a reasonable balance between expenditure on transmitter power and transmitting aerial gain. The calculation does not show whether we need 15, 20, 25 or 30 dB relative to the reference power.

For curve (i) the cost shows a minimum at 12·5 dB relative to 1 kW (just under 20 kW). If one accepts a lower standard of performance such as would be given by curve (iii), the minimum cost is achieved with a power of 6 dB relative to 1 kW (4 kW). An important factor in these calculations is the lowest frequency required to maintain service, and in the particular case shown it is 5 Mc/s.

Franklin aerials have to be adjusted to within 1% of the operating frequency, and this presents some difficulty when one has to build an aerial before one knows the operating frequency to this accuracy. I like Franklin aerials (among others), and I think all the existing ones in the United Kingdom are in use.

I am glad that the question of flexibility has been raised, and agree that it can be overdone; it is not excessive at Rugby and in the amplifiers there is very little that is not needed for simple telegraphy, although they are also excellent for telephony.

Open-wire feeder losses are small compared with those of corresponding coaxial cables.

The Radio Regulations encourage the use of the term "peak power," but I should explain, for the benefit of non-radio engineers, that it is a mean power over a short interval and the definition is set out in full in the Radio Regulations.

Mr. W. H. Lee (*communicated*): From the operational aspect the main weakness in the design of the new transmitters is the association of preset harmonic generator and mixer units with a continuously tunable r.f. amplifier. If a transmitter fails on traffic—and even the best transmitters sometimes do this—there is no spare equipment immediately available. For this reason it is suggested that two transmitters be completely motorized so that they can quickly be tuned to any frequency in the 4–27·5 Mc/s band.

The telegraph distortion produced by a h.f. transmitter is sensibly negligible; the usual source is between the terminal station and the radio station, yet the automatic monitors indicate only distortion in the transmitter itself. They are thus likely to give the central controller a false sense of security, for distorted or mutilated signals could be radiated without an alarm being given. Again, when the radiated envelope differs from the input to the telegraph drive unit, e.g. when frequency-shift diplex is used, complicated equipment is necessary to reconstitute the radiated signal for comparison purposes, and this might well produce more distortion than the negligible contribution by the transmitter.

In order to exploit radiotelegraph services to the limit it is axiomatic that distortionless signals be radiated; surely then it would have been more prudent to have installed signal regenerators at the radio station and to have guarded against signal failure with a much simpler and cheaper indicator than the automatic monitor.

Mr. H. Page (*communicated*): The permissible sharpness of the radiated beam for a point-to-point service is limited by uncertainty regarding the ionospheric path. Ignoring special cases such as the New Zealand circuit, what minimum limits would the authors suggest for the angular spread of the beam in the horizontal and vertical planes? Are the values for these parameters shown in Fig. 21 dictated by technical or by economic considerations?

It would appear from Fig. 21 that the authors favour projection angles of 5°–20° to the horizontal; but projection angles as low as 1°–2° are possible, and for these, if we assume that the ionosphere is smooth and spherical, a measure of ionospheric

focusing results. Theoretically this effect can result in an appreciable increase in received field strength over a long path, compared with propagation using higher elevation angles. Can this increase be achieved in practice? I realize, of course, that to obtain low projection angles means using much higher masts, but the advantage might be worth the greater complication.

THE AUTHORS' REPLY TO THE ABOVE DISCUSSION

Capt. C. F. Booth and **Mr. B. N. MacLarty** (*in reply*): We are glad that Mr. Stanesby emphasizes the value of good frequency stability and of low harmonic content, since these two factors are still very important to efficient spectrum usage. It is interesting to note that low harmonic radiation has been achieved by careful design of the transmitter output coupling circuits. No attempt was made to design filters to deal with specific harmonic frequencies because such filters would not be practical, having regard to the wide range of frequencies over which the transmitters are required to operate.

Mr. Faulkner rightly points out that the new station is an excellent example of co-operation between the Post Office and the manufacturer. In raising the question of frequency nomenclature, he knows that one of the authors has been a protagonist of a rationalized system;* however, the current use of the terms 'Bands I, II, III,' etc., for the higher broadcasting bands, results in some ambiguity which he wishes to avoid. The matter will be further studied by the C.C.I.R. at the VIIIth Plenary Assembly, August, 1956, and we are hopeful that a satisfactory solution will be found.

In reply to Mr. Smale, we carefully examined the merits of grouping transmitters in several unattended buildings, and concluded that the reduced feeder losses would not offset the flexibility and simplicity given by a central building. In any case, the power consumption is only part of the economic problem; the cost of the total power intake to the station is much below the figure of £100 000 he suggests for the annual loss.

Air or water cooling could have been used, but when all factors are considered air cooling has the advantage in transmitters of this type in which the power dissipated is small. The simple and clean design of the radio-frequency circuits would have been impossible with water cooling, the inevitable r.f. losses in water insulating coils would have been incurred, and the maintenance of a water-cooling system serving 28 transmitters would be considerable. We have no doubt that the decision to use air cooling was a wise one.

The reversal of the sideband signals on telephony below 10 Mc/s, as compared with operation above 10 Mc/s, has been the practice for a long time and was agreed internationally in 1953. The practice for telegraphy is different, and brings in other complications. It is necessary at Rugby to cater for both operating conditions. The reversal of the drive conditions is done automatically.

The automatic signal monitors are designed to check the operation of the radio equipment on single-channel telegraphy, and on single- and multi-channel telephony. Experience will show to what extent faults on the individual channels of a multi-channel system should be indicated at the transmitting station. As Mr. Lee suggests, the distortion produced by the transmitters is very low. Clearly, it is the first objective to achieve a high standard of performance and reliability in the main transmission chain.

* BOOTH, C. F.: Nomenclature of Frequencies', *Post Office Electrical Engineers' Journal*, 1949, 42, p. 47.

It is desirable to limit the initial switching transients on the h.v. supply circuits; the reduced-voltage condition is applied only for a few milliseconds, and any failure to reach the full voltage condition would result in a fault indication.

With regard to Mr. Smale's comments on the reliability of variable inductors, possibly he is not aware that the inductors used in these transmitters were based on a design developed in 1938 which was used in a BBC transmitter for eight years without trouble. Furthermore, a large number of medium-wave broadcasting transmitters using very large inductors of a similar type have been in operation since 1939 with entirely satisfactory results.

On the questions raised by Messrs. Thomas and Page on rhombic-aerial performance, it should be noted that some seven designs with widely differing characteristics are employed. The larger rhombics were introduced to give improved performance on frequencies below 10 Mc/s. The characteristics shown in Fig. 21 are those calculated for the quoted designs; these represent a major advance on earlier practice, but should not necessarily be taken as optimum values.

Because of the expected difficulty in working directly to Singapore about noon, the U.K.–Singapore point-to-point circuits have been operated via a relay station in Nairobi for several years.

The stacked array of dipoles for the New Zealand service was adopted to give a moderate gain by restriction of the vertical radiation pattern whilst retaining the horizontal lobe necessary to cover the very wide variations in angles of transmission.

The answers to the six questions listed by Mr. Thomas are:

(*a*) No. Normal heating is used in winter and forced ventilation in summer.
(*b*) The limits differ for different types of equipment, the maximum being about 500 watts.
(*c*) No detailed information is immediately available.
(*d*) Measurements are accurate to within 1 c/s. Observations at Baldock are made as part of the normal long-term checking procedure, but are unnecessary for day-to-day operation.
(*e*) Performance is very satisfactory and records are available.

The values for crosstalk between adjacent feeders are of the same order or better than that to be expected between nearby aerials, and the resulting intermodulation between transmissions is primarily related to the crosstalk between aerials. No *ad hoc* arrangements have been necessary.

Mr. Lee's suggestion that the association of pre-set harmonic generator and mixing units with a continuously variable r.f. amplifier results in no spare being immediately available is incomplete. Whether a spare is available will depend on the traffic schedules of the station. If one is available, any transmitter can be set up on any frequency, the motorized part of the transmitter being adjusted by manual control of the motors and without troubling to preset the automatic stopping mechanisms. The harmonic generators can be tuned by screwdriver controls. When we have two self-tuning transmitters, we may make them all self-tuning.

621.376.54 : 621.396.65

Paper No. 1900 R
Sept. 1955

PULSE-TIME-MODULATION TERMINALS FOR MUSIC TRANSMISSION OVER RADIO LINKS

By R. F. ROUS, B.Sc.

The paper was first received 17th September, 1954, and in revised form 14th April, 1955. It was published in September, 1955, and was read before the Radio and Telecommunication Section *11th January, and the* North-Western Radio Group *15th February, 1956.)*

SUMMARY

A communication system is described which enables three music circuits and an engineer's circuit to be combined. The combination is effected by using pulse time modulation in four time-division multiplexed channels. The purpose of the system is to enable a micro-wave link, intended for the transmission of television signals, to carry audio signals of broadcast-programme quality as an alternative. The radio equipment, primarily intended for television transmission, need not be specially designed for this particular service.

Measurements of the performance of the equipment indicated very satisfactory signal/noise and signal/crosstalk ratios. When the terminals were directly connected the figures obtained were better than those recommended by C.C.I.F. When the terminals were operated in conjunction with a 25-mile radio link there was a 20 dB margin in the signal/noise ratio over the recommended value. This margin was allowed for fading of the radio signal.

The choice of time division as a multiplexing method is discussed, and the selection of the pulse repetition frequency, duration and time deviation is explained.

A detailed description of the multiplexing circuit is given, and both the performance specification and actual measured performance figures are tabulated. Finally, the signal/noise ratio obtained with the radio link working over a 25-mile path is discussed, and the improvement which results from the use of pulse time modulation is explained.

(1) INTRODUCTION

Communication systems handling telephony signals of relatively narrow bandwidth have been in general use for many years. However, it should be realized that wider-band systems for music circuits are of considerable importance. These circuits are termed "normal programme circuits" by the C.C.I.F., who have made certain performance recommendations.

Music circuits capable of carrying programme material over long distances were well established 20 years ago,[1,2] but the use of multiplexing equipment to carry several programme transmissions does not seem to have occurred until much later. In 1938 C. W. and E. I. Green in the United States[3] proposed a method for using the frequency allocation normally allotted to two speech channels of a multi-channel carrier telephony system, so that a programme transmission could be multiplexed with the other telephone transmissions. A system using the r.f. band below that of a 12-channel carrier telephony system was in operation in the United Kingdom in the same year.[4]

The principle of adding one programme circuit to multi-channel telephone systems was still being developed after the Second World War,[5] but no attempt seems to have been made to multiplex several programme transmissions over a carrier system. In 1946, however, Grieg[6] suggested the use of time-division multiplex for programme transmissions, and the use of pulse time modulation was envisaged. In the next year a 24-channel system for programme multiplexing was developed, and pulse time modulation was employed.[7] This multiplexing equipment was used in conjunction with a u.h.f. beam radio link, and although satisfactory results were reported, there would appear to have been no further publication.

The equipment described in the paper employs pulse time modulation and operates with an s.h.f. beam radio link. So far as is known, the performance of this equipment is better than has hitherto been achieved by previous production equipment using pulse modulation.

(2) DESIGN CONSIDERATIONS

(2.1) Operational Requirements

The pulse-time-modulation (p.t.m.) equipment to be described was developed to allow three programme circuits and a spare or engineer's circuit to be routed over a single radio link. The equipment is intended to operate with a radio link designed for carrying television signals. The multiplexed signal is sent as an alternative to the television signal, and hence an alternative facility is provided on a radio link which is not otherwise readily adaptable. The major limitation of this type of radio link is the amplitude response of the modulation circuits, which makes the radio equipment quite unsuitable for operation with a frequency division multiplex (f.d.m.) system.

The three programme circuits of the equipment described are designed to meet a performance specification which is in all respects equal to or better than the C.C.I.F. recommendations for normal programme circuits.[8] This performance specification, which applies to the terminal equipments working back to back, ensures that the terminals contribute very little to the total noise power in the programme circuits. The expected performance with the radio link working over a 25-mile path meets the C.C.I.F. recommendations, with a suitable margin in signal/noise ratio to allow for fading.

(2.2) The Performance Specification

Most of the design considerations of the multiplex equipment are based on the performance requirements, and it is convenient to enumerate these requirements here.

The programme channels are designated channels 1, 2 and 3; the spare or engineer's channel is channel 4.

All figures are relative to a test tone level of +8 dBm, which is the maximum useful level (crash point) of a channel. The test tone has a frequency of 800 c/s unless otherwise stated.

These figures are equal to or better than those recommended by the C.C.I.F., the principal improvements being those of signal/noise ratio and harmonic distortion. The C.C.I.F. recommend a weighted signal/noise ratio of 56 dB relative to the maximum level of the channel. The specification figure of 65 dB ensures that the terminal equipment will contribute negligible noise power when the radio equipment is working under limit conditions, e.g. when severe fading occurs. The C.C.I.F. further recommend provisionally that the total harmonic distortion of a programme signal at maximum level should not exceed −20 dB relative level. This figure was considered unacceptable, and a

The paper is a communication from the Staff of the Research Laboratories of The General Electric Co., Limited, Wembley, England.

figure of −40 dB for the relative level of individual harmonics is specified.

	Channels 1, 2 and 3	Channel 4
Crosstalk from any other channel	−74 dB or better	−58 dB or better
Unweighted noise ..	−36 dB or better	−20 dB or better
Weighted noise (1949 C.C.I.F. music weighting)	−65 dB or better	−49 dB or better
Bandwidth..	50 c/s–10 kc/s	50 c/s–10 kc/s
Harmonic distortion measured at 800 c/s ..	Individual harmonics −40 dB or better	Individual harmonics −40 dB or better
Harmonic distortion measured at 100 c/s ..	Individual harmonics −35 dB or better	Individual harmonics −35 dB or better
Variation in overall gain with time	±2·0 dB within 24 hours	±2·0 dB within 24 hours

(2.3) The Multiplexing Method

The main problem in the design of the multiplexing equipment was to find a multiplex signal suitable for transmission over a radio link of relatively poor amplitude linearity. A time-division multiplex signal in which pulse time modulation is employed was eventually chosen in preference to other methods as the best solution to the problem.

The first multiplexing method considered was a frequency-division system in which four amplitude-modulated sub-carriers are combined. This method was rejected because of crosstalk limitations. Crosstalk will occur in the channels of a frequency-division system when the combined signal is subject to amplitude distortion. Cross-modulation and harmonic-distortion components will in general give rise to non-intelligible crosstalk which will have the same effect as added noise. It was found that a considerable improvement in the amplitude linearity of the radio modulation circuits was required if the performance specification figure of −65 dB for weighted noise was to be satisfied. For this reason the use of frequency-division multiplex was not attempted.

A multiplexing method in which a sub-carrier is frequency modulated by a frequency-division signal was also considered.[9] This scheme was not attempted because of the complexity of the apparatus required and the engineering difficulties involved. The chief difficulty here was the design of modulation and demodulation circuits of adequate linearity for the frequency-division signal.

Time-division multiplex with pulse time modulation was the method eventually chosen, and the reasons for this choice may be summarized as follows:

(*a*) The modulated pulses are position-modulated (time-modulated) and distortion of the pulse amplitude is unimportant.

(*b*) The frequency and phase characteristics of the radio circuits were suitable for a pulse-time-modulation signal. (Intelligible crosstalk generally arises with a typical pulse-time-modulation signal when phase shifts of the components of the modulated pulse spectrum are severe.)

(*c*) Although a pulse-time-modulation signal requires a wide base band for transmission, this band is available with a television-type radio link. The disadvantage of the relatively low signal/noise ratio obtained with a wide-band system is offset to some extent by the improvement obtained with pulse time modulation.

(2.4) The Pulse Characteristics

When pulse time modulation is employed there are three main characteristics of a time-modulated pulse train that must be considered, namely

(*a*) The pulse repetition frequency.
(*b*) The pulse shape.
(*c*) The time deviation of the pulse at maximum modulation level.

Fig. 1 shows the multiplexed pulse train; the dotted lines about channel 1 pulse show the time-deviation limits when the channel is fully loaded (test-tone level). The repetition frequency employed is 32 kc/s per channel, which gives a repetition period of 31·2 microsec for the combined pulse train. The time allocation for each of the programme channels is 7·8 microsec, the remaining 7·8 microsec period being shared between the synchronizing pulse and channel 4.

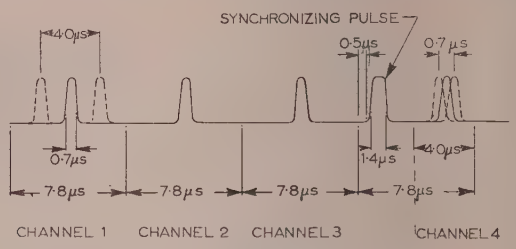


Fig. 1.—The multiplexed pulse train.

(2.4.1) The Pulse Repetition Frequency.

The repetition frequency employed is 32 kc/s per channel. This ensures that the programme signals are sampled at a rate which just exceeds three times the maximum modulation frequency encountered (10 kc/s). Although a sampling rate which just exceeds twice this frequency is the basic minimum with a pulse modulation system, the higher value has been chosen so that certain distortion terms introduced by the modulation process are made to lie outside the audio-frequency band.

Briefly, the modulation process is one in which the sampling instants and the pulse occurrence times are synchronous. Demodulation is effected by a conversion to pulse width modulation followed by a low-pass filtering. With such a process the principal distortion components that remain after demodulation are multiple sidebands of the repetition frequency.[10] If a repetition frequency of twice the maximum modulation frequency is employed, and the pulse duration and time deviation are appropriate to a 4-channel system, only two major distortion components are obtained within the audio-frequency band. These occur at frequencies $f_r - 2f_m$ and $f_r - 3f_m$, where f_r is the repetition frequency and f_m is the modulation frequency. The relative levels of these components are −25 dB and −45 dB respectively; these figures are not acceptable. The higher value chosen for the repetition frequency removes the sideband at $f_r - 2f_m$ from the audio-frequency band, and hence a distortion level of better than −40 dB is achieved.

An alternative modulation process which was considered involved periodic sampling of the modulating signal at the transmitter and conversion to pulse amplitude modulation at the receiver.[11] This method does not give rise to the distortion mentioned above, and a lower repetition frequency is permissible. However, the additional equipment required for this method is quite considerable, and an increase in the repetition frequency was considered the best practical solution to the distortion problem.

(2.4.2) **The Pulse Shape.**

The pulse train which is passed by the p.t.m. transmitter to the radio equipment is shown in Fig. 1. This results from the application of rectangular pulses of 0·7 microsec duration to a low-pass filter having a cut-off frequency of 2 Mc/s. The shape of the pulse train is not changed by the radio equipment, and the pulse train which arrives at the input terminals of the p.t.m. receiver is the same as that shown in Fig. 1, but with added noise.

At the input terminals of the p.t.m. receiver the pulses pass through a low-pass filter which has a cut-off frequency of 1·5 Mc/s. The resulting pulses have a raised-cosine shape with added noise. This raised-cosine shape was found by experiment to give the best compromise between signal/noise and signal/crosstalk ratios. The duration of the pulses at half-height was also chosen on this basis.

When the signal/noise ratio is the only consideration it can be shown[12] that an optimum pulse duration exists. If the pulse duration at half-height is doubled, the minimum transmission band is halved. Consequently, the total noise power present in this band is reduced by 9 dB when frequency modulation is employed in the radio link. This improvement results from the familiar triangular noise spectrum that is obtained with frequency modulation. Furthermore, the pulse slope at half-height will be halved, and the improvement in the signal/noise ratio which is obtained with pulse time modulation is reduced by 6 dB. The total effect of doubling the pulse duration is therefore to raise the signal/noise ratio in the audio circuit by 3 dB, provided that the time deviation of the pulse is unchanged. However, the final signal/noise ratio depends directly on the time deviation of the pulse. The fraction of the repetition period allocated to a channel is fixed, and this time interval must be shared by the pulse duration and the pulse time deviation. The pulse duration which gives the best signal/noise ratio must therefore compromise between these two considerations, and it is deduced to be 1·5 microsec for the system.

It was found by experiment that the crosstalk due to the trailing edge of a 1·5 microsec pulse was unacceptable. A value of 0·7 microsec duration was eventually chosen, although this involved a sacrifice of some 3 dB in the signal/noise ratio.

(2.4.3) **The Pulse Time Deviation.**

The peak-to-peak time deviation of the programme channel pulses is 4·0 microsec at maximum modulation level (+8 dBm). This is approximately half of the total time allocation per channel, which is 7·8 microsec; hence a margin is allowed for setting up errors and overload signals. Tests on the double-diode limiter which is employed in the modulator circuit as a safeguard against crosstalk (see Section 3.1.2) showed that an overload signal 20 dB above test-tone level was peak-clipped so that the resulting pulse time deviation used only 70% of the 7·8 microsec allocation. The peak-to-peak time deviation of channel 4 pulse is 0·7 microsec at maximum level. This ensures that the synchronizing pulse which appears within the same 7·8 microsec period is protected. Since a lower time modulation depth is employed for this channel, the signal/crosstalk and signal/noise ratios for this channel are reduced relative to those for the programme channels. In fact, a degradation of 16 dB in both ratios is suffered.

(3) THE TERMINAL EQUIPMENT CIRCUITS

The complete communication link for the four channels is shown in block diagram form in Fig. 2. It can be seen from the diagram that the programme multiplexing equipment may be conveniently divided into two main units: the transmit terminal and the receive terminal. For this reason these units will be described individually.

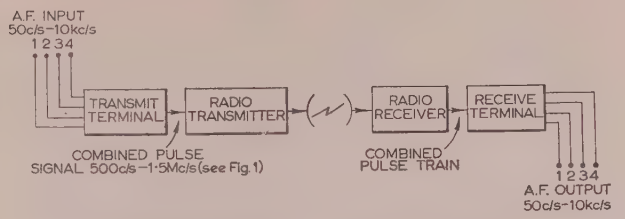


Fig. 2.—Block diagram of the 4-channel link.

(3.1) The Transmit Terminal

The operation of the terminal is best explained with reference to Fig. 3. A common waveform generator provides four trains of timing pulses for the operation of the channel modulators.

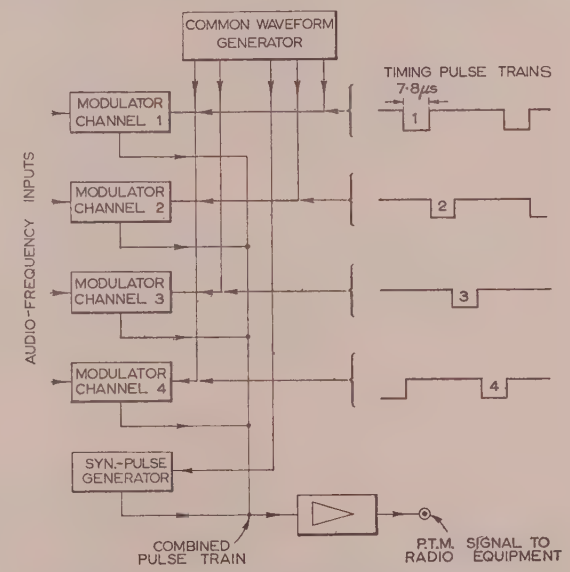


Fig. 3.—Block diagram of the transmit terminal.

The four pulse trains are interlaced in a repetition period of 31·2 microsec, the repetition frequency of each being 32 kc/s. The timing pulses are of 7·8 microsec duration and define the pulse modulation periods allotted to the channels. Each of the four audio-frequency signals enters a modulator unit, where with the aid of the timing pulses a train of time-modulated pulses is obtained. The four modulated pulse trains thus produced and a synchronizing-pulse train are added and applied to a common amplifier. The synchronizing-pulse train is produced with the aid of a further train of timing pulses supplied by the common waveform generator. These timing pulses are synchronous with those that operate channel 3 modulator and are not shown in the diagram. The synchronizing pulses are timed by the trailing edges of the timing pulses and thus occur just within the 7·8 microsec periods allotted to channel 4. The pulses are of 1·4 microsec duration, i.e. twice the duration of the modulated pulses. The combined pulse signal is shaped and amplified in the common amplifier and appears finally as a positive-going pulse train of 1-volt amplitude in an unbalanced impedance of 75 ohms. This output is convenient for conveying the signal to the radio equipment by coaxial line.

(3.1.1) The Common Waveform Generator.

The common waveform generator provides all the timing waveforms for the operation of the transmit terminal. A crystal-controlled master-oscillator produces a 32 kc/s sine wave. This is applied to a phase-shift network which provides four sine-wave outputs in quadrature. Each sine wave is combined with one of two 64-kc/s square waves derived from the oscillator waveform. The combined waveforms are shaped and amplified by four output stages which provide the four timing-pulse trains for the operation of the channel modulators. The output stage which supplies channel 3 timing waveform also provides an auxiliary output for the operation of the synchronizing pulse generator. The circuits employed in this unit are quite conventional, and no circuit diagram is given.

(3.1.2) The Modulator.

The channel-modulator circuit diagram and its associated waveform diagrams are shown in Fig. 4. The function of the modulator is to produce a train of short time-modulated pulses with the aid of the timing waveform and the audio-frequency signal.

The audio signal entering the unit is applied to a low-pass filter with a cut-off frequency of 10 kc/s. The filtered signal is then applied to a step-up transformer, T_1, which has a double-diode limiter circuit connected across its secondary winding. The two diodes V_1 are biased so that peak-clipping of the audio signal occurs when the input level to the modulator exceeds +8 dBm.

The audio signal is then applied to one grid of a double-triode comparator, V2, the function of which stage is to produce a train of pulses (iii) which are width-modulated by the audio signal. A sawtooth waveform (ii) produced by integration of the timing pulse (i) is applied to the other grid of this valve. The operation of the stage is as follows:

The triodes V_{2A} and V_{2B} are biased so that in the absence of signal on either grid the former is cut off and the latter is conducting. When the sawtooth excursion lowers the grid potential of V_{2B} to some voltage near that of the grid of V_{2A}, current begins to flow in V_{2A}, and as the negative excursion at the grid of V_{2B} continues this valve is cut off. Since the sawtooth excursion covers many times the grid base of either valve, the change-over in the flow of anode current is rapid and the pulse of anode current in V_{2A} is steep-sided. When the sawtooth flyback causes the grid voltage of V_{2B} to go positive again, this valve is rendered conducting, V_{2A} is cut off, and the pulse of anode current in V_{2A} is terminated. Now the grid voltage of V_{2A} is controlled by the audio signal during the switching process, and the voltage on the grid of V_{2B} varies linearly with time. The instant at which both voltages differ by some fixed amount which allows anode-

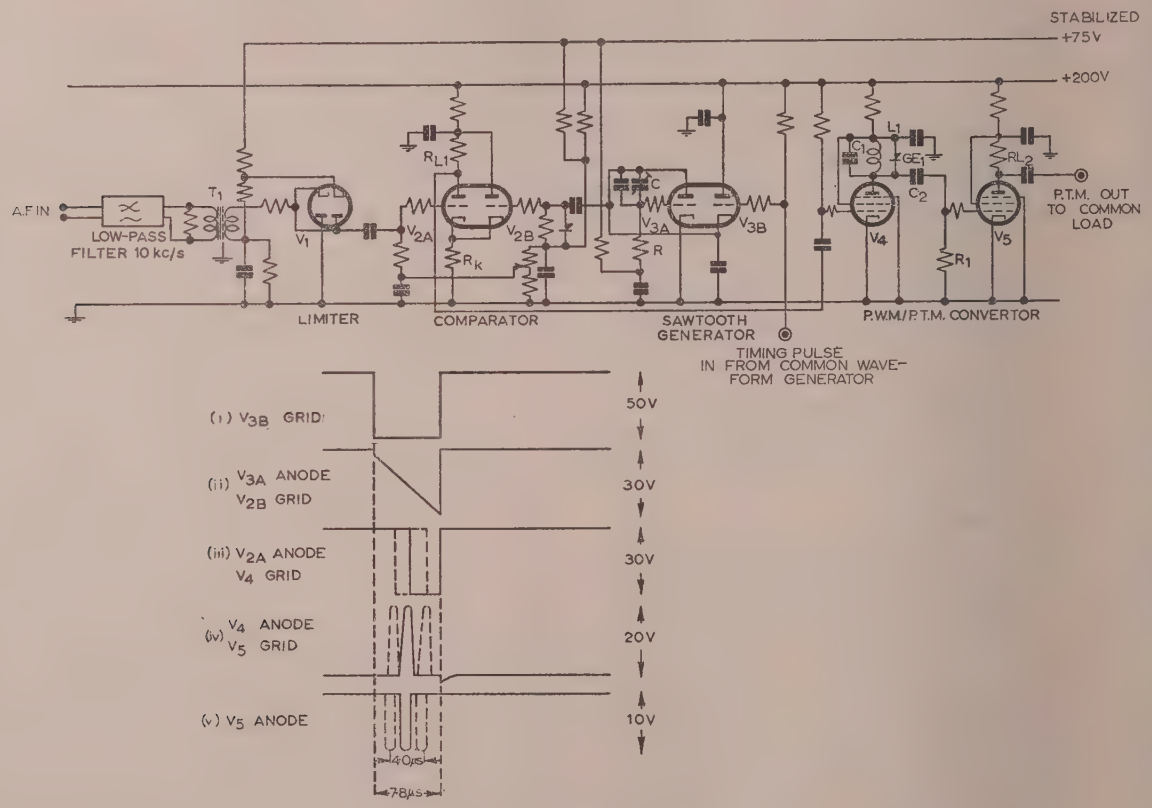


Fig. 4.—The channel modulator.

current flow in V_{2A} to begin is therefore dependent on the value of the audio-signal voltage at this time. The leading edges of the pulses of anode current in V_{2A} are thus time-modulated by the audio signal. Since the trailing edges of these pulses are fixed by the flyback times of the sawtooth waveform, the pulses are width-modulated by the audio signal.

In order to limit the effect of changes in valve characteristics over the working voltage range of the comparator, a large common cathode load R_k is provided. The grid resistors of either triode are returned to stabilized supply voltage points. This arrangement ensures that the grid-voltage condition that just allows anode current to flow in V_{2A} is almost independent

the magnitude of the audio signal and depends solely on the fference between the two grid voltages. In this way the stortion introduced by the non-linearity of the valve charac-ristics is reduced to an acceptable level.

The integration process by which a linear sawtooth waveform produced with the aid of the timing pulse train is carried out y a Miller integrator valve V_3. The voltage excursion produced y this process is, in fact, a part of an exponential decay of time-onstant μCR, where μ is the amplification factor of the valve $_{3A}$, C is the feedback capacitance and R is the grid resistance.

if they were allowed to appear in the combined pulse train, are removed by the class-C action of the amplifier stage V_5.

The pulse signal (v) which appears at the anode of V_5 is negative-going and is passed to a common point where the pulse signals from the other modulators appear.

(3.1.3) The Synchronizing-Pulse Generator and Common Amplifier.

The synchronizing-pulse generator and the common amplifier are combined into a single unit, the circuit and waveforms for which are shown in Fig. 5. The unit produces a train of short

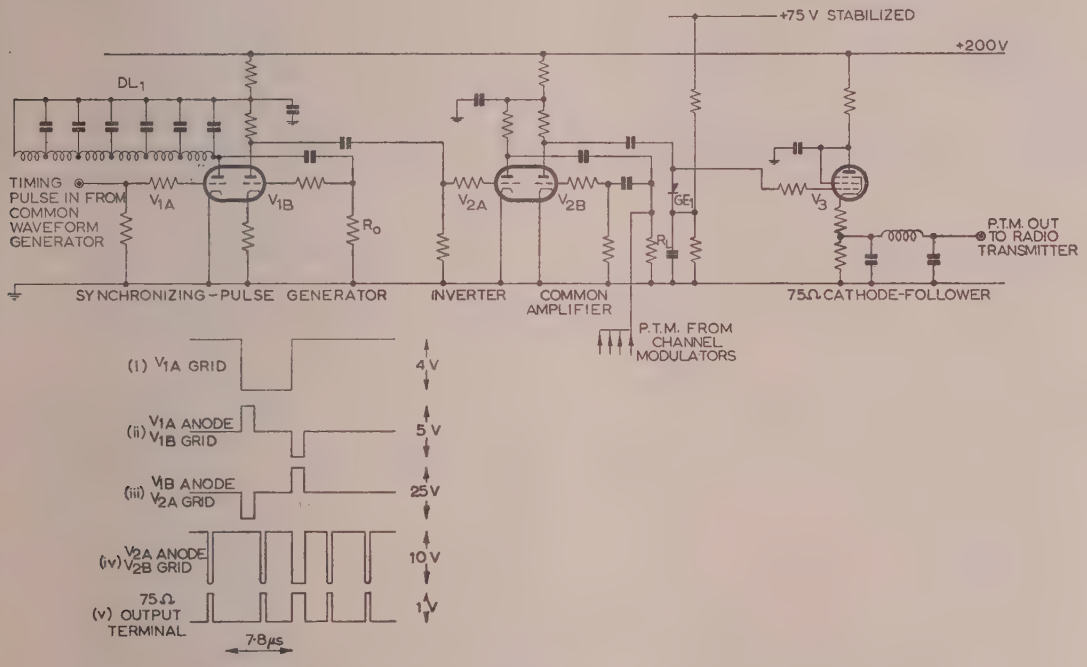


Fig. 5.—The synchronizing-pulse generator and common amplifier.

These values are arranged so that a time-constant of approximately 1 millisec is obtained. This ensures that the voltage excursion is virtually linear during the 7·8 microsec period of the sawtooth, and no appreciable distortion is introduced.

The width-modulated pulse train which appears across the comparator valve anode load of RL_1 is applied to grid of the convertor valve V_4. This valve draws grid current during the positive excursion of the waveform, and is cut off during the negative excursion, so that peak-clipping of the width-modulated pulses is achieved. The rise-time of the pulses is improved considerably by this process, and traces of audio signal introduced by phase shifts in the anode decoupling circuit of V_2 are reduced. The anode-current pulses of V_4 are used to shock-excite a tuned circuit L_1, C_1, which has a germanium diode GE_1 connected in parallel. The leading edges of the width-modulated pulses thus give rise to short positive pulses (iv) which are time-modulated, while the trailing edges of the wide pulses are rendered ineffective by the action of the diode.

The time-modulated pulse train which appears at the anode of V_4 is applied to the grid of a class-C amplifier. The low-frequency response of the coupling network C_2, R_1 is deliberately degraded to reduce the amplitude of audio components in the pulse signal, introduced by phase shifts in the decoupling network in the anode-voltage supply line for V_4. Residual traces of this signal, which would give rise to crosstalk pulses which are used for synchronization purposes, combines them with the four modulated pulse trains from the channel modulators and shapes the combined pulse train.

A train of timing pulses from the common waveform generator is applied to the grid of the synchronizing-pulse generator valve V_{1A}; this pulse train is negative-going. The pulses of anode current which flow in V_{1A} are used to drive a delay line DL_1, which has a short-circuit termination. The delay of this line is 1·4 microsec (forward and return), and consequently a reflected current pulse appears at the line input terminals with a delay of 1·4 microsec relative to the initial driving pulse. A resistance R_0 is connected to the input terminals of the line and acts as a termination, so that further reflections do not occur. The reflected current pulse combines with the input pulse to the line to produce the waveform (ii) shown in Fig. 5. This waveform is applied to the grid of V_{1B}, which acts as a class-A amplifier. The waveform (iii) which appears at the anode of V_{1B} has the leading edges of its positive excursions synchronous with the trailing edges of the original timing pulses. The pulses (iii) are applied to the grid of a class-C amplifier V_{2A}, where only the positive-going pulses are amplified. The negative-going train of 1·4 microsec pulses which appears on the anode of V_{2A} is then suitable for use as a synchronizing-pulse train, to be combined with the 4 channel pulse trains.

The four interlaced time-modulated pulse trains from the

channel modulators are applied to a common load resistor R_L; the synchronizing pulses from the anode of V_{2A} are also connected to this load. The combined pulse train is then applied to the grid of a common amplifier V_{2B}. All the pulses of this train are able to cut the valve off, and consequently the resulting anode-current pulses are of equal magnitude. The combined pulse train which appears at the anode of V_{2B} is about 30 volts in amplitude and is positive-going. The positive peaks of this pulse train are restored to a point at +10 volts by current flow in a germanium diode, GE_1. The restored waveform is then applied directly to the grid of a cathode-follower V_3. Since the negative excursions of the restored waveform lower the grid voltage to −20 volts, the valve is cut off during these excursions. This process combined with the d.c. restoration ensures that the current pulses which flow in the cathode load of V_3 are of constant amplitude, and normal variations in the valve characteristics are unimportant. The cathode load of V_3 is tapped to provide a low output impedance, and a low-pass filter is connected there so that the video signal is restricted to the band 0–2·0 Mc/s.

(3.2) The Receive Terminal

Fig. 6 shows a block diagram of the receive terminal. The received pulse train is applied to a common unit where the

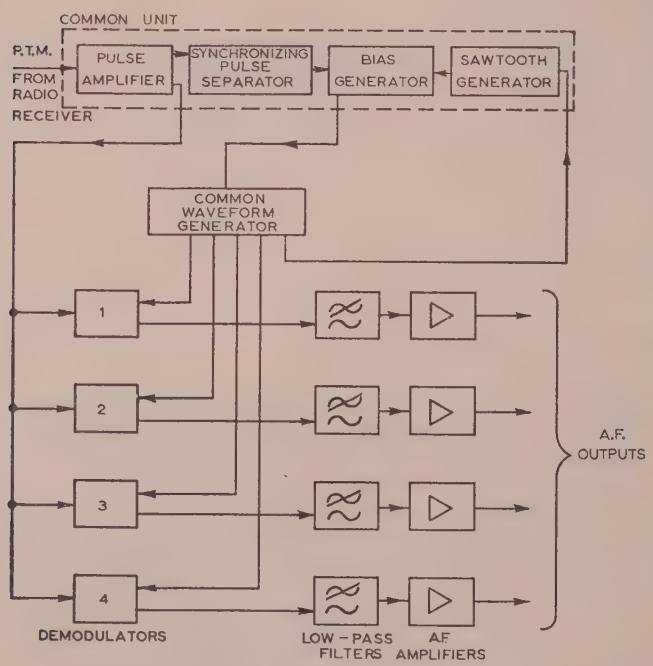


Fig. 6.—Block diagram of the receive terminal.

pulses are amplified and sliced, and then passed to the channel demodulators. This unit also contains the circuit for separation of the synchronizing pulses which operate an electronic servo-mechanism, which in turn controls the receiver master oscillator in another unit—the common waveform generator. This generator performs a similar function to the common waveform generator of the transmit terminal. Four interlaced trains of timing pulses are produced; the frequency and phase of each pulse train is determined by the master oscillator, which operates at 32 kc/s. Each timing-pulse train is applied to a channel demodulator unit, where it is used to facilitate the separation of the appropriate train of time-modulated pulses. The separated pulse train is then used to generate a train of width-modulated pulses which are applied to a low-pass filter. This filter passes only the audio components of the pulse-width-modulation signal, and thus the original modulating signal is recovered. The four audio signals from the demodulators are passed to four a.f. amplifiers which provide the final outputs of the terminal.

(3.2.1) The Common Waveform Generator.

The operation of the receive-terminal common waveform generator is similar to that in the transmit terminal. There is one major difference, however, in that the master oscillator is not crystal-controlled but is locked in frequency and phase by an electronic servo-mechanism known as a "spongy lock." This spongy lock which forms part of the next unit to be described provides a control bias which is applied to a reactance valve. This valve forms part of the main tuned circuit of the oscillator.

As in the transmit terminal, the waveform generator supplies four trains of timing pulses to the channel units—in this case to the demodulators. A further train of timing pulses is supplied to the sawtooth generator which drives the spongy lock.

(3.2.2) The Synchronizing-Pulse Separator and Bias Generator.

Amplifier and slicer stages shape the received pulse train, pass it to the channel units, and also to the synchronizing-pulse separator stage. A circuit diagram of this stage and other associated circuits is shown in Fig. 7. The synchronizing-pulse amplifier V_1 is driven by a negative-going train of the received pulses. The anode-current pulses of V_1 are then used to drive a delay line DL_1 which has an open-circuit termination. The total delay of the line (forward and return) is 1·0 microsec, and after this time a reflected wave from the termination combines with the initial pulse train to form the waveform (i). Double-amplitude pulses are formed from the synchronizing pulses, which are of 1·4 microsec duration, but not from the channel pulses, which are of 0·7 microsec duration. The synchronizing pulses are then separated from the double pulse train in a class-C amplifier V_{2A}. This valve conducts during the peak excursions of the double-amplitude synchronizing pulses and is cut off for the remainder of the time. The anode-current pulses which flow during this operation are shown as waveform (ii). These current pulses are applied to a transformer T_1, which has a double-diode switching circuit connected across its secondary winding. A sawtooth generator V_{2A} supplies the sawtooth waveform (v) to the switch at A. The two diodes V_3 are rendered conducting during the separated synchronizing-pulse period, and current flows from A into a storage capacitor C_1. The voltage developed across this capacitor is passed to the common waveform generator to control the reactance valve.

The sawtooth waveform (v) is produced by integrating a pulse train (iv). This pulse train appears at the termination of a delay line DL_2, which is supplied by the waveform generator with a timing pulse train (iii). The time delay of the line DL_2 is such that the sawtooth waveform is in the centre of its steep excursion when the separated synchronizing pulses (ii) occur. This condition is, of course, maintained only when the receive terminal is correctly locked to the incoming signal. A potentiometer VR_1 enables the mean level of the bias voltage developed across C_1 to be varied. The relative phases of the received pulse train and the receiver oscillator waveform may then be adjusted in the locked condition. A change in the setting of VR1 causes the sawtooth to be switched at a different time, i.e. at a different voltage level, so that constant bias is maintained, and the locked condition is held. Alternatively, temperature changes in the oscillator which would vary the frequency are rendered ineffective by slight variations in phase which result from the servo action. In this way changes of ±300 c/s in the

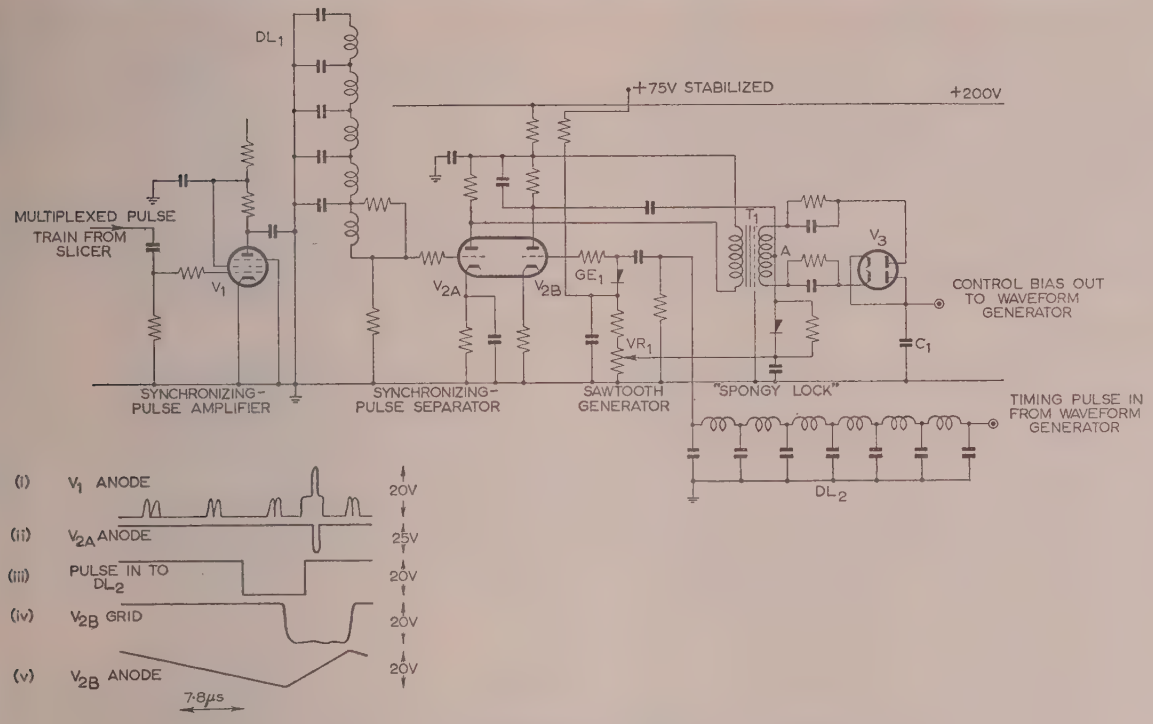


Fig. 7.—The synchronizing-pulse separator and bias generator.

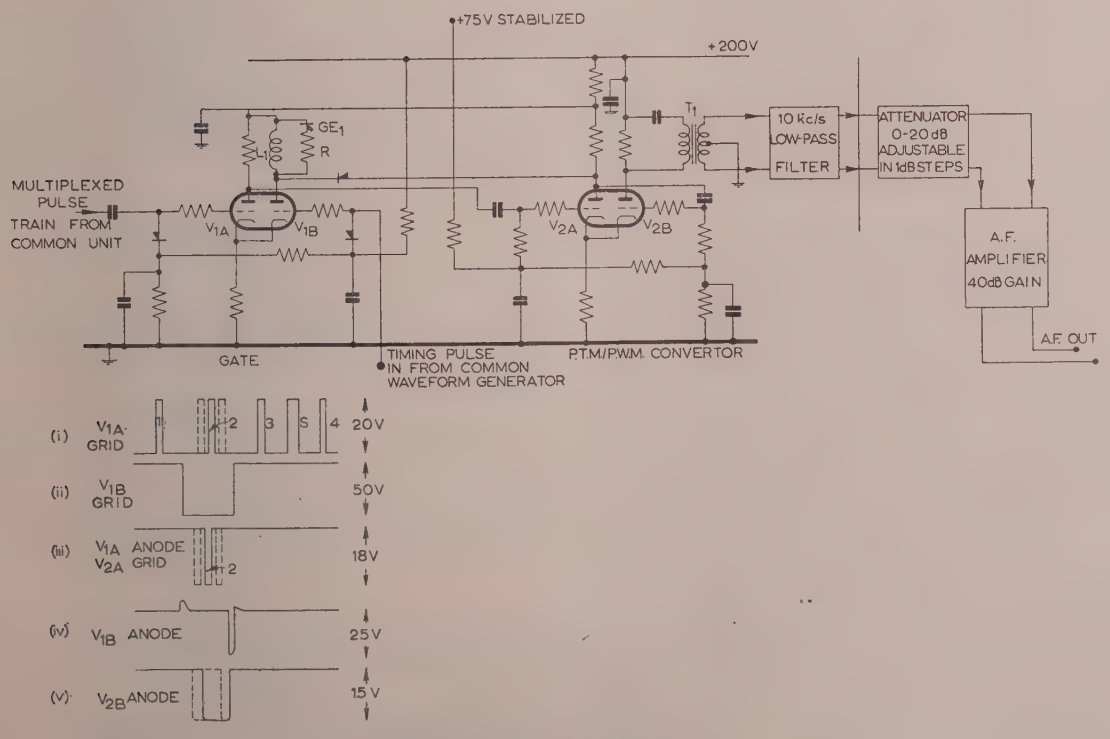


Fig. 8.—The channel demodulator.

nominal frequency of the oscillator may be nullified. The change in the relative phase which results is $\pm 1 \cdot 0$ microsec, and this is the "pull in" range of the servo mechanism.

The main problem encountered in the design of this unit was one of crosstalk reduction. Stray signals fed back from the synchronizing-pulse-separator delay line DL_1 introduced severe crosstalk, and a pentode amplifier was found to be necessary to drive the line. Further improvements in crosstalk were obtained by changes in component layout, and separation of wiring in the amplifier and slicer stages. In some circuits, extra decoupling capacitors were found to be necessary, so that phase shifts of the low-frequency components of the p.t.m. signal were reduced.

(3.2.3) The Demodulator.

The demodulator is designed to select the appropriate channel pulses from the received pulse train, and to produce a width-modulated pulse train which may be filtered to give the original audio signal. A circuit diagram and the principal waveforms are shown in Fig. 8.

The demodulator is supplied with the received pulse train which has been sliced and amplified; this pulse train (i) is applied to one grid of a double-triode V1. The other grid of this valve is supplied with a train of timing pulses (ii) from the common waveform generator. The bias-voltage supplies to which the two grid resistors are returned are arranged to keep V_{1B} conducting and V_{1A} cut off in the absence of signal at either grid. When a negative-going pulse from the timing train lowers the grid voltage of V_{1B}, this valve is cut off. V_{1A} is then able to conduct during the positive excursions of the multiplexed waveform, but is cut off during the base-line excursion of the waveform (i). This valve remains cut off until a further channel pulse and timing pulse are coincident. The separated train of channel pulses is obtained by this process, and appears as waveform (iii) at the anode of V_{1A}. The anode-current pulses in V_{1B} which result from the negative-going timing pulses at the grid of V_{1B} are used to shock-excite a tuned circuit consisting of an inductance L_1 and stray capacitances. This tuned circuit is shunted by a germanium diode GE_1 in series with a resistor R. The diode conducts during each current pulse and the circuit is heavily damped by R. At the termination of each current pulse the diode is momentarily cut off and a short pulse is developed across the tuned circuit. During the next excursion of this oscillation the diode again conducts and the oscillation is damped out.

The separated channel pulses (iii) and the short pulses (iv) derived from the trailing edges of the timing pulses are used to trigger a multivibrator stage which contains a double triode V_2. Triode V_{2A} is normally conducting and is cut off by each negative-going channel pulse. The change of voltage at the anode of V_{2A} then switches V_{2B} from a cut-off to a conducting condition. V_{2B} is again cut off when a negative-going pulse of waveform (iv) triggers back the multivibrator. The pulses of anode current which flow in V_{2B} have their leading edges defined by the channel pulses, which are time-modulated, and their trailing edges defined by the pulses (iv), which are regularly spaced. The anode-current pulses are thus width-modulated by the original audio signal. This signal is separated from the other components of the width-modulated pulse spectrum by the action of a low-pass filter which has a cut-off frequency of 10 kc/s. The filtered audio signal is then passed to an amplifier, which provides the output signal of the channel.

To ensure constant amplitude of the width-modulated pulses, stabilized grid voltage is supplied to V_2. In this way the amplitude of the audio output signal is virtually independent of the demodulator action, and small changes in valve characteristics are rendered unimportant.

(4) THE PERFORMANCE OF THE EQUIPMENT

(4.1) Crosstalk

Crosstalk measurements at 800 c/s showed that signal/crosstalk ratios were better than 80 dB for all combinations of programme circuits. These measurements were carried out with a tone at the maximum useful level of a programme channel (+8 dBm). The very-low-amplitude crosstalk signals were found to change in level very rapidly as the positions of the channel pulses were varied. This pointed to the presence of high-frequency oscillations following each pulse, presumably due to imperfect transient response or to stray signal feedback from delay-line circuits. The effect of simple phase changes due to inadequate frequency response could not be detected.

(4.2) Noise

Signal/noise ratio measurements with the terminals directly connected gave ratios of 75 dB weighted (1949 C.C.I.F. music weighting) and 56 dB unweighted; both measurements were relative to the maximum useful level of the channel. The effect of weighting the noise should be to reduce the signal/noise ratio by 5 dB if the noise is white. This was not the case, and it must be assumed that most of the unweighted noise was due to power-supply hum, since the C.C.I.F. weighting factor for a 100 c/s tone is $-26 \cdot 1$ dB.

(4.3) Distortion

Harmonic distortion tests with 800 and 100 c/s tones at maximum level showed that individual distortion components were less than 1% of the fundamental. Measurements with $7 \cdot 5$ and $8 \cdot 5$ kc/s test tones to determine the unharmonic distortion (see Section 2.4.1) showed the largest distortion component to be -45 dB relative to the fundamental.

(4.4) Level Stability

The variation in the overall gain of a channel was measured over a period of 24 hours, during which the terminal equipment was supplied with constant mains voltage. The variation was found to be $\pm 0 \cdot 3$ dB. A further test with $\pm 10\%$ variations in the mains supply voltage showed the maximum overall variation in level to be $\pm 0 \cdot 7$ dB. This meets the C.C.I.F. requirement of $\pm 2 \cdot 0$ dB, but it should be realized that this recommendation applies to a 1 000 km link.

(5) THE SIGNAL/NOISE RATIO

The signal/noise ratio in a programme channel may be determined if the signal/noise ratio in the video circuit of the radio receiver is known. The calculation must take into account the improvement in signal/noise ratio that is obtained by the use of pulse time modulation. This improvement is discussed in Section 8.

(5.1) The Unweighted Signal/Noise Ratio

The performance data for the radio equipment show the expected signal/noise ratio after transmission over a 25-mile path to be

$$20 \log_{10} \left(\frac{\text{peak-to-peak signal in video circuit}}{\text{r.m.s. noise voltage in 3 Mc/s baseband}} \right) = 59 \cdot 1 \text{ decibels} \quad . \quad . \quad . \quad . \quad (1)$$

The p.t.m. signal requires a baseband of only $1 \cdot 5$ Mc/s for transmission of the $0 \cdot 7$ microsec pulses, and so

$$20 \log_{10} \left(\frac{\text{peak pulse amplitude}}{\text{r.m.s. noise voltage in } 1 \cdot 5 \text{ Mc/s baseband}} \right) = 68 \cdot 1 \text{ decibels} \quad . \quad . \quad . \quad . \quad (2)$$

he improvement of 9 dB of eqn. (2) over eqn. (1) results from e triangular noise spectrum of a frequency-modulation system. urthermore, it is shown in eqn. (11) of Section 8 that the provement obtained with pulse time modulation is

$$0 \log_{10} \left(\frac{\text{r.m.s. signal/noise ratio in audio circuit}}{\text{peak pulse amplitude/r.m.s. noise in baseband}} \right) = 10 \text{ decibels} \quad . \quad . \quad . \quad . \quad (3)$$

should be realized that this improvement is relative to the gnal/noise ratio given in eqn. (2), which deals with the peak-to-eak video signal. If, then, r.m.s. values are considered through-ut, the total improvement is seen to be 19 dB.

From eqns. (2) and (3) it follows that

$$0 \log_{10} \left(\frac{\text{r.m.s. signal voltage in audio circuit}}{\text{r.m.s. noise voltage in audio circuit}} \right) = 78{\cdot}1 \text{ decibels} \quad . \quad . \quad . \quad . \quad (4)$$

Now the audio circuit has an available bandwidth of 16 kc/s, ecause a pulse repetition frequency of 32 kc/s is employed. ut for a programme circuit, only 10 kc/s of this bandwidth is ctually used (see Section 2.4.1), so that the overall signal/noise atio (unweighted) in the programme circuit is found to be

$$0 \log_{10} \left(\frac{\text{r.m.s. signal voltage}}{\text{r.m.s. noise voltage}} \right) = 78{\cdot}1 + 10 \log_{10} \left(\frac{16}{10} \right) = 80{\cdot}1 \text{ decibels} \quad . \quad (5)$$

(5.2) The C.C.I.F. Weighting

Although the noise distribution in the video circuit follows he familiar triangular spectrum associated with frequency nodulation, it does not follow that such a distribution will arise n the channel circuits. Individual noise components in the ideo band will behave as sidebands of the pulse-repetition-requency harmonic which is nearest in frequency, and the ctual channel noise will have a fairly uniform distribution. This means that the effect of the C.C.I.F. weighting charac-eristic may be determined on the basis of a white-noise distri-ution in the channel. The result obtained is that the weighted signal/noise ratio in a channel will be 5·2 dB less than the flat signal/noise ratio. The C.C.I.F. recommendation for weighted noise is that a ratio of 56 dB should be obtained. This means hat a flat signal/noise ratio of 61·2 dB is required. The calculated figure of 80·1 dB allows a margin of 19·9 dB, which was con-sidered an adequate fading allowance.

(6) CONCLUSIONS

The equipment described in the paper is probably the first production equipment which can multiplex several programme signals of broadcast quality so that the combined signal is suitable for transmission over a radio link. The use of pulse time modula-tion ensures that the radio link need not have an extremely linear amplitude response; a link designed for the transmission of television signals is generally acceptable.

Measurements on the multiplexing equipment have shown that performance figures equal to or better than those recom-mended by the C.C.I.F. for normal programme circuits, can be obtained with pulse time modulation.

(7) REFERENCES

(1) Clark, A. B., and Green, C. W.: "A Long Distance Cable Circuit for Program Transmission," *Bell System Technical Journal*, 1930, **9**, p. 567.

(2) Hamilton, H. S.: "A Wide Band Open Wire Program System," *Electrical Engineering* (New York), 1934, **53**, p. 550.

(3) Green, C. W., and Green, E. I.: "A Carrier Telephone System for Toll Cables," *Bell System Technical Journal*, 1938, **17**, p. 101.

(4) Halsey, R. J., and Tucker, D. G.: "The Provision of Music Circuits on Twelve Channel Carrier Cables," *Post Office Electrical Engineers' Journal*, 1938, **31**, p. 280.

(5) Leconte, R. A., Penick, D. B., Schramm, C. W., and Wien, A. J.: "A Carrier System for 8,000 cycle Program Transmission," *Bell System Technical Journal*, 1949, **28**, p. 165.

(6) Grieg, D. D.: "Multiplex Broadcasting," *Electrical Communication* (New York), 1946, **23**, p. 19.

(7) Altman, F., and Dyer, J. H.: "Multiplex Broadcasting," *Electrical Engineering* (New York), 1947, **66**, p. 372.

(8) C.C.I.F. Yellow Book 1951, Tome III bis., p. 154.

(9) Gerlach, G. C.: "A Microwave Relay System," *RCA Review*, 1946, **17**, p. 576.

(10) Fitch, E.: "The Spectrum of Modulated Pulses," *Journal I.E.E.*, 1947, **94**, Part IIIA, p. 556.

(11) Parks, G. H., and Moss, S. H.: "A New Method of Wide Band Modulation of Pulses," *ibid.*, p. 511.

(12) Feldman, C. B., and Bennett, W. R.: "Bandwidth and Transmission Performance," *Bell System Technical Journal*, 1949, **38**, p. 40.

(8) APPENDIX

(8.1) The Signal/Noise Ratio Improvement

When a train of time-modulated pulses is subject to added noise, this noise disturbs the positions of the pulse edges. The ratio of the pulse time deviation due to the modulating signal and the time disturbance due to noise gives the signal/noise ratio in the audio circuit.

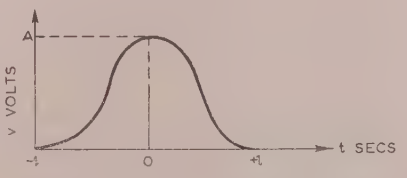


Fig. 9.—The raised-cosine pulse.

The raised-cosine pulse shown in Fig. 9 may be expressed as

$$v = \frac{A}{2}\left(1 + \cos \frac{\pi t}{l}\right) \quad -l < t < l \quad . \quad . \quad . \quad (6)$$

The rate of change of pulse amplitude is thus

$$\frac{dv}{dt} = -\frac{A\pi}{2l} \sin \frac{\pi t}{l} \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

which has a maximum value of $\pm A\pi/(2l)$ when $t = \pm \frac{1}{2}l$. The r.m.s. time disturbance in the position of the pulse edge at the point of maximum slope is therefore

$$\sqrt{(\overline{\Delta t^2})} = \sigma \frac{2l}{A\pi} \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where σ is the r.m.s. noise voltage in the video band. If the peak time deviation due to the modulating signal is T, the signal/noise ratio in the audio circuit will be

$$\frac{\text{r.m.s. noise voltage in audio circuit}}{\text{r.m.s. signal voltage in audio circuit}} = \frac{T}{\sqrt{2}\sqrt{(\overline{\Delta t^2})}} = \frac{A\pi T}{(2\sqrt{2})l\sigma} \quad . \quad . \quad . \quad . \quad (9)$$

from eqn. (8). But A/σ is simply the signal/noise ratio in the video circuit. The improvement in signal/noise ratio is then

$$\frac{\pi T}{(2\sqrt{2})l} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

Reference to Section 2.4 for the values of T and l gives

Peak time deviation $T = 2{\cdot}0$ microsec.
Pulse duration at half-height $l = 0{\cdot}7$ microsec.

Thus

$$\text{signal/noise improvement} = 20\log_{10}\left(\frac{2\pi}{2\sqrt{(2)} \times 0{\cdot}7}\right)$$
$$= 10 \text{ decibels} \quad . \quad . \quad . \quad . \quad (11)$$

DISCUSSION BEFORE THE RADIO AND TELECOMMUNICATION SECTION, 11TH JANUARY, 1956

Dr. A. R. A. Rendall: Occasions arise when it is very advantageous to use one of our existing s.h.f. television links to transmit ordinary sound programmes. Normally, of course, no one would propose doing such a thing, because of the very wide frequency range absorbed. I emphasize 'existing s.h.f. links', which were designed for television, and whose linearity is inadequate for multiplexing using ordinary means of modulation. If we had been free to choose, the decision might have been to improve the radio link rather than to adopt pulse-time modulation. The whole matter must be viewed in the light of our particular request to use existing apparatus for a special purpose.

The apparatus is inevitably rather complicated, and it is interesting to recall that when we contemplated planning our v.h.f. programme transmissions for the Home, Light and Third programmes, suggestions were made that we might multiplex them so that one transmitter could radiate three programmes and a suitable receiver could deal with them. The paper, to some extent, answers the question why we did not attempt to do it that way. For one thing, the frequency band employed would be excessive, and secondly, if we tried to economize in frequency band we should be faced with non-linearity and intermodulation. Thirdly, the receivers would have been rather complex.

The requirements of the s.h.f. link might, I think, be rather severe in certain respects, and I would ask the author whether he has considered the effect of phase distortion on the s.h.f. link itself. If this system were used over a long s.h.f. link there might be phase distortion, which can be distributed over the frequency band, although of course it is more at the upper end of the band. With the pulses used, there may be an overshoot due to phase distortion which would cause crosstalk.

I suppose that this is the reason why a raised-cosine pulse was used, because with that particular pulse the ripples resulting from phase distortion are likely to be small. From a noise point of view I imagine a pulse with much sharper sides would be preferable.

I was a little dismayed to note that peak clipping has been resorted to. Over the years in the B.B.C. we have had long arguments about what is the right way of limiting a programme signal. We frequently have to do it. In the case of our v.h.f. transmissions we do in fact limit it in order to avoid excess deviation. In our view, peak clipping is unsatisfactory, as it introduces distortion. We have instead adopted the method of variable gain control, and when a peak which will overload comes in, the gain of the amplifier is depressed quickly and the recovery time is rather slow.

The great advantage of a peak chopper is, I suppose, simplicity. In programmes excess amplitude is bound to occur on occasions. This particular link might be tied to a long land-line circuit into which there were switched various sections, and when we switch and re-switch our land-line connections we suffer considerable variations in amplitude. It is essential to take care of these, as the author has said, and I wonder whether this is done in the best way.

When discussing distortion and 'unharmonic distortion', the author chooses test-tones of 7·5 and 8·5 kc/s. I wonder why those tones were chosen. They are not very representative of what happens in programme itself, because, as is well known, frequency components of that order are very weak in programme, and those of major amplitude occur much lower in the frequency band. I am glad the author did not take the C.C.I.F. figure for distortion of −20 dB. I think that all broadcast engineers realize that that figure dates back to a long time ago and was put in because we did not know of anything better. The B.B.C. circuits meet a much better specification than that; in fact, we normally keep to about 1% or better. We weight the harmonic components according to where they occur in the frequency band, because from an annoyance point of view the effects of various harmonics are not equally harmful.

One rather surprising omission from the paper is that no overall frequency characteristic is given for the apparatus when operating over a s.h.f. link. We can assume that the amplitude characteristic of the s.h.f. link is blameless, but I wonder whether other things which might arise in ordinary transmissions would affect the amplitude characteristic.

I have raised these points because I want to emphasize the complexity of the problem which was presented to the manufacturers of this equipment. We realized that it was a difficult assignment. We tried to do it ourselves by simple modulating methods, and failed, for the reasons which the author gives. He has tackled the problem thoroughly and has dealt with the various difficulties. The solution which is offered to us seems to be eminently satisfactory.

Mr. C. W. Earp: While acknowledging that the development described has produced a satisfactory result, it is of interest to examine where communication efficiency could theoretically have been improved.

First, presumably owing to the choice of a system in which signal sampling rate is not uniform, distortion has been avoided by unusually generous margins of safety with regard to the pulse group-repetition period and the amount of pulse-time deviation. The theoretical maximum group-repetition period for 10 kc/s fidelity is, of course, 50 microsec, which would permit each of 4 channel pulses to have a deviation of about ±6 microsec. In practice, in order to avoid distortion and crosstalk, a figure of ±2 microsec has been chosen, so that a theoretical deficiency of 10 dB arises.

Secondly, I would draw attention to the receiver used, which in the demodulation process utilizes one edge only of each pulse. As a rough approximation, an ideal receiver should give a 3 dB improvement in signal/noise ratio as compared with the method chosen for the practical case.

Thirdly, I would mention the theoretically possible use of the instantaneous compandor. There is not, perhaps, one which is perfect enough for broadcast quality, but such a device can offer another 20 dB in effective signal/noise ratio, without penalty of additional bandwidth.

Fourthly, I would draw attention to the radio-frequency head, where the video pulse train, which has an on/off ratio of

nly 1 : 10, is used to frequency-modulate a final carrier wave. The carrier wave is thus undisturbed for nine-tenths of the time, which must represent some waste of power. My approximate stimate is that four-fifths of the transmission is effectively nused, representing a further deficiency of 7 dB.

The deficiencies already mentioned add up to a total of 40 dB, which does seem rather large.

To be practical, and if possible more constructive, I admit that know of no really satisfactory compandor, but I would like to ut forward the case for pulse-amplitude-modulated (p.a.m.) requency modulation. In a 4-channel p.a.m. base-wave, where ignal sampling takes place at a uniform rate, a group time-base f 40 microsec should be satisfactory, and of 10 microsec avail-able for each pulse, effective use could be made of at least 5 microsec as compared with 4 microsec for the pulse-position-modulated (p.p.m.) case. Further, the demodulation of p.a.m. frequency modulation is more nearly ideal than that of single-edge p.p.m. demodulation. Also, there is no obvious deficiency regarding incomplete modulation of the carrier wave. I there-fore suggest that a practical development of the p.a.m. frequency-modulation system might well show an overall improvement factor of about 10 dB.

Mr. A. P. Carter: The relative simplicity of the p.t.m. appara-tus in comparison with, for example, a currently used 1 + 4 f.d.m. telephony terminal is apparent from the circuit details given, though there is bound to be some difference of opinion about the exchange of the passive filter elements of the f.d.m. system for the electronic elements of the p.t.m. method. In bulk, the p.t.m. terminal occupies about the same bay-space as the 1 + 4 f.d.m. system. Each unit of the p.t.m. terminal may be rapidly withdrawn for servicing, and comprehensive metering facilities are provided. The alignment of the terminals is relatively simple in practice, and the accurate positioning of the pulses can be readily checked with a commercially available double-beam oscillograph in association with the X-sweep provided from the common waveform generator.

The use of p.t.m. instead of sub-carrier frequency modulation with 4 f.d.m. channels, mentioned in Section 2.3, would appear to have been a wise choice from the operational point of view, although the facility of integrating the channel group with the existing f.d.m. carrier system is lost.

It would be interesting to hear the author's views on employ-ing the present system with the channel bandwidth modified to carry, say, a 12-channel telephony group, as this could permit the off-loading of certain telephone traffic on to a television radio circuit in the event of emergency or heavy commitments.

In Section 2.4.2, reference is made to the triangular noise spectrum of a f.m. system and to the 9 dB reduction in noise level when the video bandwidth is halved. Measurements made on three different systems employing crystal and valve limiter combinations indicate a reduction of about 6 dB instead of 9 dB. It would be interesting to know whether the author has observed a similar divergence from the theoretical value.

In Sections 2.1 and 5 it is claimed that the C.C.I.F. weighted signal/noise ratio of 56 dB is met on a certain 25-mile radio circuit. However, when the system was tested over a 20-mile radio circuit, the video signal/noise ratio approximated to that quoted in Section 5.1, but the weighted noise level was 5 dB above the C.C.I.F. value when the s.h.f. input to the receiver was attenuated by 20 dB. In terms of the author's 25-mile radio circuit, it would appear that a practical figure for the fading margin is approximately 13 dB.

Mr. W. Ross: We hear a good deal about communication over fairly long distances by scatter propagation, and I should like to know whether this type of modulation (p.t.m.) is particularly appropriate for that form of communication.

Mr. L. Stenning: The question of a compandor has been raised. It is quite true that if we measure a signal/noise ratio in a certain manner a compandor can be used to increase the signal/noise ratio. In other words, if we measure a peak signal and then take it off and measure the noise, and call that ratio the signal/noise ratio, then it is susceptible of improvement by a compandor. However, we can also measure the signal/noise ratio simultaneously, and under those conditions the compandor does not give the same improvement of the ratio.* Also, I think it would be extremely difficult to achieve less than 1% distortion with a compandor giving ten times the slope at the origin compared with the linear system; i.e. it would be diffi-cult to maintain the proper inverse relation between the trans-mitting compressor and the receiving expander. That diffi-culty, however, is not theoretical; it is merely one of practical engineering.

Some apprehension has been expressed about the use of instantaneous limiting. Something has to be done to restrict the audio wave within certain limits, or the modulation of the radio pulse will go outside its proper range and interfere with the next pulse. A constant-volume amplifier can, of course, be used in addition.

Mr. Earp said there is a 10 dB loss because the sampling rate was 30 kc/s instead of 20 kc/s. That is actually a 5 dB loss, I think. I do not wish to infringe on the author's prerogative of replying to these points, but I do not believe that a p.a.m. system could possibly achieve the quality required: the crosstalk due to low phase distortion at the low-frequency end would be completely prohibitive. I should think it would be difficult to achieve a crosstalk value much better than 40 dB with the p.a.m. system.

Dr. A. R. A. Rendall: Mr. Stenning compared a peak chopper with a quick-acting limiter, but what I said was that the limiter was fast in the uptake but slow in restoration.

THE AUTHOR'S REPLY TO THE ABOVE DISCUSSION

Mr. R. F. Rous (*in reply*): The requirements of the s.h.f. link are not particularly severe for a p.t.m. signal, but as Dr. Rendall suggests, the raised-cosine pulses employed minimize the effects of phase distortion. It must be emphasized that it was not intended to use several radio equipments to give a long link, and consequently the effect of only one link was considered. Any undesirable ripples added to the pulse edges by phase distortion at the h.f. end of the video band in the modulation circuit will be rejected by the low-pass filter in the video circuit of the p.t.m. receiver. It is not true to say that a steeper pulse edge would give an improved signal/noise ratio. This point is dealt with in the paper, where the optimum value of the video bandwidth is discussed.

The use of peak clipping in the programme circuits has been criticized, and the use of variable-gain amplifiers was suggested. However, a sharp restriction on the time-modulation depth of the pulses must be imposed to prevent the severe crosstalk which can arise when pulses of one channel break into the time allocation of an adjacent channel. There is, of course, no objection to the use of variable gain-control circuits in the lines supplying the programme signals to the p.t.m. transmitter.

* MALLINCKRODT, C. O.: *Bell System Technical Journal*, 1951, **30**, p. 706.

The test-tone frequencies of 7·5 and 8·5 kc/s were chosen in order that the anharmonic distortion terms equal to the repetition frequency minus three times the modulation frequency fell within the 0–10 kc/s band of the programme channel. Other tests at lower frequencies were, of course, carried out to determine the harmonic distortion.

Mr. Earp suggested that a repetition frequency of 20 kc/s should be employed, and that a time deviation of ±6 microsec would then be permissible. This time-deviation value does not allow for the duration of the pulse and assumes that the peak-clipping circuits are perfect. A more realistic value for the time deviation with a 20 kc/s repetition rate would be ±3 microsec. The small improvement in signal/noise ratio thus obtained does not warrant the added complexity of the periodic sampling equipment which would be required.

The use of a double-edge demodulation process was rejected on the grounds that, at best, a 3 dB improvement in signal/noise ratio is obtained, and this does not justify the use of additional equipment. Furthermore, the stated improvement is obtained only when the video noise spectrum is white, and with the triangular noise spectrum obtained with an f.m. system, the improvement would probably be negligible.

The suggested use of an instantaneous compandor has been answered by Mr. Stenning, who stated that, even if the desired overall linearity could be achieved, unpleasant effects would arise. Thus, in the presence of a loud low-frequency signal, noise would arise at the peak excursions, owing to the action of the receiver component of the compandor.

The improvement in signal/noise ratio which Mr. Earp suggests would be obtained with a pulse-amplitude-modulation system is certainly desirable. Unfortunately, the amplitude linearity specified for television equipment is quite unsuitable for the programme signals, since a maximum distortion of more than 1% is permitted. Furthermore, if the video band were restricted to the basic minimum for the p.a.m. signal, in order to obtain the best signal/noise ratio, some difficulty would be experienced with the crosstalk associated with long decay periods of the pulse edges. Any attempt to reduce this crosstalk by the use of steeper pulses would, of course, result in a degradation of the signal/noise ratio.

Mr. Carter suggested that the p.t.m. terminals could perhaps be modified to carry a 12-channel telephony group. This could be achieved if the repetition frequency were increased slightly and if some improvements in the linearity of the p.t.m. modulator circuits were obtained. Both of these modifications are perfectly possible, and should be relatively simple to carry out. The discrepancy in the figure for the signal/noise ratio with an f.m. system was certainly not observed during our development work on the equipment, but the expected margin of 20 dB was not obtained during field trials. This was thought to be due to the presence of obstacles in the radio path.

621.396.11 : 621.317.34.029.6

Paper No. 2011 R
May 1956

DIRECTIONAL OBSERVATIONS ON H.F. TRANSMISSIONS OVER 2100 KM

By E. N. BRAMLEY, M.Sc., Ph.D.

(*The paper was first received 1st November, and in revised form 29th November*, 1955.)

SUMMARY

Using a wide-aperture spaced-loop direction-finder, measurements have been made of the direction of arrival of high-frequency radio waves propagated over 2100 km. Pulse transmissions were used for most of the experiments, and first- and second-order F-reflections could usually be identified at 11 Mc/s. The bearing fluctuations of these echoes included a lateral deviation component of approximately the magnitude expected from previous experiments at shorter distances. The rapid fluctuations were appreciably larger than they were at 700 km, and this appears to be associated with an increased complexity of the individual pulses. This effect was very marked in the night-time observations on 5 Mc/s and resulted in a standard deviation of 1·5° for individual bearings in an hourly period. The corresponding figure for the 1F echo in the day-time on 11 Mc/s was only 0·6°.

The results were unaffected by such changes in the transmitting aerial as were practicable, or by ionospheric or magnetic storms.

(1) INTRODUCTION

In recent years a programme of directional observations has been carried out on short waves reflected from the ionosphere, at distances up to 700 km, and the results have been interpreted in terms of ionospheric irregularities.[1,2] The present paper describes an extension of this work to a greater distance of transmission, namely 2120 km. The measurements were made to determine the inaccuracies in bearings at this distance caused by ionospheric effects, as compared with the results obtained at shorter distances.

(2) DESCRIPTION OF EXPERIMENTS

The measurements to be described were carried out between December, 1952, and October, 1954, on signals transmitted from Malta, the observations being made with a wide-aperture spaced-loop direction-finding system[3] at Winkfield, Berkshire. This direction-finder consisted of two pairs of aerials, each with 100 m spacing, one pair being spaced in the north–south and the other in the east–west direction. The direction of arrival was determined from the two phase differences between the signals in each pair of aerials.

The transmissions were provided by the Royal Air Force, whose very effective co-operation is gratefully acknowledged. For the main part of the programme the signals were pulse modulated, a pulsing unit being used in conjunction with a standard transmitter, type SWB8. The pulse width was 150 microsec, and the repetition frequency was 50 pulses/sec, derived by frequency division from the 100 kc/s crystal oscillator of a Loran indicator. A similar crystal oscillator and chain of frequency dividers were used at the receiver to control a linear time-base on which the pulse signals were displayed. Any desired pulse could be selected for study by means of the usual arrangement of an adjustable strobe applied to the grids of the cathode-ray tubes displaying the phase information.

The measurements were made photographically at 30 sec intervals, and the observations were generally carried out over periods of two hours or more at a time. Both day and night observations were made, but in the summer very few day-time measurements were possible owing to inadequate signal strength. For day-time working a frequency of 11·21 Mc/s was used, the transmitting aerial being an elevated horizontal half-wave dipole. Two such dipoles were available, oriented in different directions, and in the course of the work a comparison of results using these two aerials was made. The day frequency often continued to be useful for some hours after sunset, but for most of the night-time work a frequency of 5·11 Mc/s was used. In this case the transmissions were from either a horizontal half-wave dipole or a vertical quarter-wave aerial.

(3) THE RECEIVED PULSE PATTERNS

(3.1) 11 Mc/s

During the day the duration of the pulse pattern rarely exceeded 1 millisec, and in the winter a recognizable series of echoes was usually obtained, as shown in Fig. 1(*a*). The most regularly

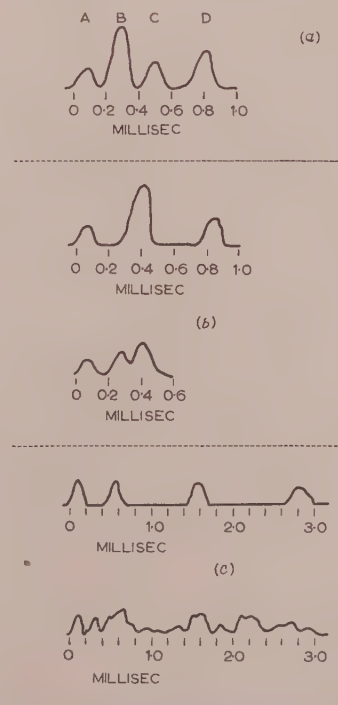


Fig. 1.—Typical echo patterns.
(*a*) Winter day; 11 Mc/s.
(*b*) Summer day; 11 Mc/s.
(*c*) Night, 5 Mc/s.

recurring components, B and D, could be identified as the first- and second-order F-reflections. The latter echo was not observed at all times, since the ionospheric critical frequency was sometimes too low to permit this mode of propagation.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

The paper is an official communication from the Radio Research Station, Department of Scientific and Industrial Research.

Ionospheric data obtained at Schwarzenburg (about 350 km from the mid-point of the transmission path) were used to assist in checking the component echoes present at any time; satisfactory agreement was obtained between the expected and observed times for 2F propagation, and provided a check on the identity of this echo. Its delay with respect to the 1F echo varied on different occasions from about 400 to 800 microsec, corresponding to equivalent heights of reflection from 210 to 310 km. The majority of the day-time measurements were made on these two echoes.

The first echo in the pattern, A, is believed to be a first-order E-reflection; it was generally very much weaker than B, and was frequently not seen at all. When present, its time of arrival was about 200 microsec before the echo B; this is the approximate interval expected between 1E and 1F echoes. The small amplitude of the 1E echo is not surprising in view of its low angle of elevation (calculated to be less than 2°), since very little energy could be radiated at such a low angle by the horizontal transmitting aerial, owing to the presence of the ground.

The echo C was also not a permanent feature of the pattern, and was rarely comparable in amplitude with the 1F and 2F echoes. This echo could be explained as an M-type (i.e. a double reflection at the F-layer with intermediate reflection from the top of the E-layer) or an N-type (i.e. involving successive reflections at the E- and F-layers, with an intermediate ground reflection). It was not, however, possible to distinguish with certainty between these two modes of propagation. The position of C was variable between B and D, so that on occasions it was not completely separated from one or other of these.

During the summer months (April–September) the pulse patterns observed in day-time rarely corresponded to the simple structure just described. This was undoubtedly due to the E-layer ionization density being too great (at any rate near noon) to allow the normal 1F mode of propagation to exist; the first significant echo received was then probably the second-order E-reflection. Fig. 1(*b*) shows two sample echo patterns observed under these conditions, but no particular pattern could be regarded as typical, and the components were often insufficiently stable for an echo to be identified over any length of time. The received signals were also weaker than in the winter, so that very few bearing measurements were made in day-time during summer. These were mainly confined to the strongest component, which almost always occurred within the first 500 microsec of the pulse pattern. With the decay of E-layer ionization at sunset, the complicated day-time pattern gave way to simpler conditions in which the 1F echo could be recognized as the main mode of propagation.

The characteristic effect associated with the maximum usable frequency, or skip, condition, namely the coalescing of high- and low-angle components as this condition is reached, was observed at all seasons, but the separate occurrence of the effect on the ordinary and extraordinary magneto-ionic components was never clearly recognized. About 5–10 min would be expected to elapse between the occurrence of the m.u.f. condition for the ordinary and extraordinary rays.

When comparisons were made between the two dipole transmitting aerials, no difference in the echo patterns, attributable to the different aerials, was observed.

A feature of the echo patterns at all times of year was that the individual pulses were only rarely as 'clean' and stable as those observed on similar frequencies at shorter distances of transmission; fading and partial splitting were generally much more noticeable.

(3.2) 5 Mc/s

At 5 Mc/s, which was used only at night, a simple echo pattern consisting of first- and higher-order F-echoes, as shown in the upper part of Fig. 1(*c*), was obtained only rarely. In general, a long train of echoes was observed, extending over 2–4 millisec, and often with no discrete pulses which retained their identity for more than a few minutes, owing to the continual fading and mixing of the components [Fig. 1(*c*), lower part]. This state of affairs was presumably due to the presence of multiple E-layer reflections along with the F-echoes. The E-echoes would be absent on the higher frequency of 11 Mc/s. Under these conditions of complex echoes, selection was carried out on a basis of the delay time from the first visible part of the A-display pattern.

(4) GENERAL DESCRIPTION OF RESULTS

Fig. 2 shows an example of the results obtained at 11 Mc/s in the day-time; individual 'snap' bearings taken on the 1F and 2F

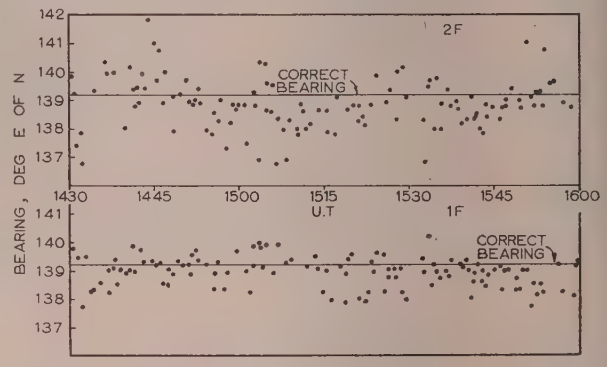


Fig. 2.—1F and 2F bearings on 11 Mc/s, 28th October, 1953.

echoes over a period of 1½ hours are shown. It is seen that for both echoes the average measured bearing is close to the correct bearing (139·2° E of N); the fluctuations about the mean are within about ±1° for the 1F and ±2° for the 2F echo.

The characteristic feature of the results of measurements at shorter distances of transmission[1,4] was the existence of both rapid and slow fluctuations (periods of the order of seconds and tens of minutes, respectively) which could be easily separated by inspection. In the present results the two types of fluctuation can again be recognized, but the rapid variations are relatively larger, so that the slow component is often observed, and, in general, no clear separation can be made. An effect frequently noted at 11 Mc/s, however, was a relative diminution of the magnitude of these rapid changes after sunset, so that the long-period bearing changes could then be more easily observed. An example is shown in Fig. 3.

At 5 Mc/s (used at night) the variations in bearing were much larger than at 11 Mc/s; an example is shown in Fig. 4. Much of the increase in spread could be attributed to rapid fluctuations arising from wave-interference effects, although this could not be regarded as due solely to 'coning' of a single mode since, as mentioned above, intermode interference is an important factor at this frequency. Occasionally, however, there were also well-marked long-period variations in bearing. The largest of these were associated with sunrise, when bearing deviations up to 10° were recorded. These were always towards the east, which is the sense to be expected from consideration of the ionospheric height changes occurring at sunrise, and this agrees with previous observations. The effective ionospheric tilt at the time could not be estimated accurately, since the mode of propagation was not known with certainty, but by making reasonable assumptions as to the identity of the echoes, taking into account the measured angles of elevation, it was found that the results were consistent with the usual 2° or 3° sunrise tilt of the F-layer. A well-marked sunset effect (in the opposite sense) was recorded on only one occasion (see Section 5.2).

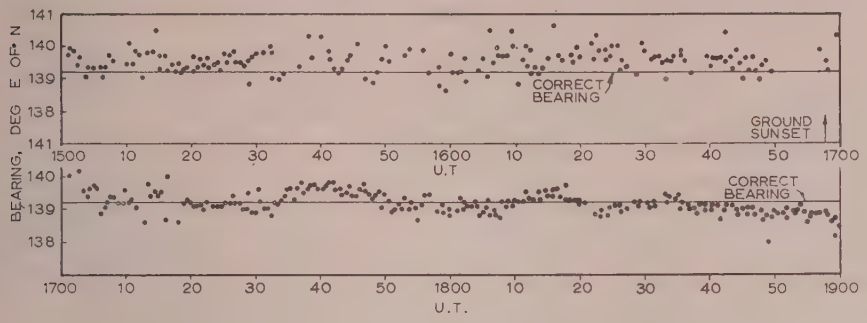


Fig. 3.—1F bearings on 11 Mc/s, 16th February, 1954.

A notable case of a large slow deviation unconnected with sunrise occurred on the 11th August, 1954, when between 0020 and 0150 U.T. the bearings on two separate components in the echo pattern both swung to the west and then returned. The deviation was about 3° on the echo with the shorter delay, and about 10° on that with the longer delay. A transverse ionospheric tilt of a few degrees, substantially uniform over a distance of several hundred kilometres along the transmission path, would account for this result.

With the spaced-aerial type of direction-finder used for the experiments, the measurements enabled the angle of elevation of the incident wave, as well as its bearing, to be determined. It is known, however, that with this type of aerial system the resolving power in the vertical plane, and hence the accuracy of elevation measurement, falls as the angle of elevation decreases. For angles less than 20°, errors up to several degrees in magnitude, with both systematic and randomly varying components, can be expected. The angles of elevation involved in these experiments were mainly below 20° (the average value for a 1F echo was about 10°) and so no great accuracy could be expected for these measurements. The values obtained, however, were generally consistent with the delay times and were of assistance in identifying echo components.

(5) ANALYSIS OF RESULTS

From each pair of phase measurements recorded during a series of observations, the instantaneous bearing of the particular echo component selected at the time was calculated, and these bearings were then plotted against time, as in the examples shown in Figs. 2–4. Because of the comparative unreliability of the angles of elevation, these were calculated only as 5 min averages, i.e. the phase readings from the two orthogonal pairs of aerials were separately averaged over periods of 5 min, and then combined to give the angle of elevation.

For a statistical analysis of the results they were divided into a number of groups, according to the echo concerned. On 11 Mc/s three different echoes were identified and observed sufficiently frequently to be analysed separately. These were the first- and second-order echoes (1F and 2F) and the first-order high-angle or Pedersen component (1F HA), which was frequently observed at times near the m.u.f. condition. On 5 Mc/s, since individual echoes could scarcely ever be identified with certainty, the results were arbitrarily classified in two groups (I and II), namely those with delays less than, and greater than, ½ millisec from the first visible part of the echo pattern. Group II consisted mainly of echoes with delays between ½ and 1 millisec from the leading edge, for, although the pattern frequently extended over 2 or 3 millisec, a sufficiently stable component with a delay of more than 1 millisec was rarely obtained.

The results in each group were divided into periods of about an hour, and for each period the mean bearing,* α_2, and the standard deviation, σ, of the individuals about this mean were calculated. In order to study the relative amounts of 'rapid' and 'slow' fluctuation occurring at any time, each period was further divided into 5 min intervals, and the mean bearing, α_1, in each interval containing at least five observations was calculated. An estimate of the standard deviation, σ_1, of the 5 min means within each hourly period was then made from the range and number of these means.[5]

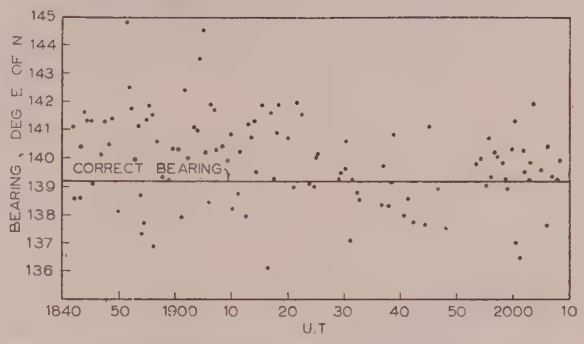


Fig. 4.—5 Mc/s bearings, first main echo, 19th January, 1954.
200–300 microsec from start of echo pattern.

(5.1) Temporal Variations

An examination of the results was made for systematic temporal variations. For each group of data, values of α_2, σ and σ_1 were plotted against time of day and time of year. In most of the plots no systematic effects were apparent, but in the 11 Mc/s 1F results there was a tendency for α_2 to fall at night and σ_1 to rise, as compared with the day values. To check the reality of this effect these results were divided into three sections, namely day, twilight and night. The division between day and twilight was taken as $\chi = 90°$, and between twilight and night as $\chi = 102°$; χ is here the sun's zenith angle at the (ground) mid-point of the transmission path. For each section the results were pooled, and the grand mean bearing, α_3, and the overall variance, v'_1, of the 5 min means about the period means were calculated, together with the estimated standard error of these quantities. Table 1 shows the results.

The twilight results are clearly more closely related to the day than the night values, and so the day and twilight results have been pooled, giving the figures shown in the last line of the Table. The term 'day' will subsequently be taken to include the whole period for which $\chi < 102°$. A comparison between the last two lines of the Table shows that the difference in α_3 between day and night is not significant at the 5% level, but the difference in

* α denotes an individual bearing, α_1 a 5 min mean, α_2 a period mean and α_3 the grand mean.

Table 1

DIURNAL CHARACTERISTICS OF 1F ECHO

	Number of bearings	Number of 5 min means	Number of periods	α_3	v_1'
				deg	deg^2
Day	4 874	548	79	139·4 ± 0·05	0·090 ± 0·007
Twilight	1 635	169	27	139·3 ± 0·10	0·11 ± 0·02
Night	1 119	126	16	139·1 ± 0·16	0·22 ± 0·04
Day and twilight	6 509	717	106	139·4 ± 0·05	0·095 ± 0·007

v_1 is significant at this level, i.e. the spread of the 5 min means about an hourly mean is greater by night than by day.

In the 1F day results on 11 Mc/s taken from September, 1953, to April, 1954, there was a suggestion of a seasonal change in the mean bearing. The mean bearings over four 2-month periods are shown in Table 2, together with their estimated standard errors.

Table 2

SEASONAL CHARACTERISTICS OF 1F ECHO

Period	Mean bearing
	deg
September, October, 1953 ..	138·6 ± 0·10
November, December, 1953 ..	139·3 ± 0·05
January, February, 1954 ..	139·6 ± 0·07
March, April, 1954	139·4 ± 0·08

These results thus indicate a significant bearing variation over the eight months, the bearings being higher in winter than in autumn and spring. Results in the corresponding months of the previous year are consistent with this behaviour, but there are insufficient summer results to enable the form of a complete annual variation to be determined. The cause of this seasonal variation is not clear. There does not appear to be any seasonal change in ionospheric structure which would suggest such a bearing variation. The possibility cannot be ruled out that the effect is an instrumental one, caused by a seasonal variation in line-up error due to ground-constant changes, but this seems unlikely.

(5.2) Transmitter Aerial Comparisons

On 11 Mc/s the two transmitting aerials were both horizontal half-wave dipoles, 26 m above ground, giving a vertical polar diagram whose major lobe extended from about 5° to 25° in elevation, with the maximum at 15°. The aerial orientations were such as to give maximum radiation in directions 284·5° and 323·5° E of N. The latter direction is close to the true bearing of the receiver from Malta (329·9°). Although the other direction was 45° away from this, the broad horizontal polar diagram of a half-wave dipole resulted in a theoretical field-strength reduction of only 35% in this direction, compared with the maximum.

The aerial comparison tests were made by observing transmissions from the two aerials during alternate half-hour periods, and the signal strength ratio was consistent with the calculated value. Four such tests were made, each extending over three hours, and in no case could any difference, either in the echo pattern or in the directional results, arising from the different transmitting aerials, be detected. Fig. 5 shows an example of the results of these tests, in which both the 1F and 2F echoes were recorded. It is seen that no change in the bearing spread, or its mean value, was caused by the aerial changes. The different

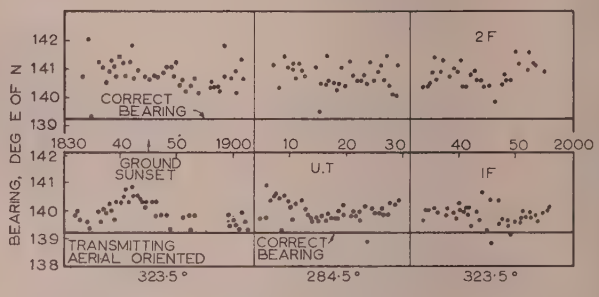


Fig. 5.—1F and 2F bearings on 11 Mc/s, 14th May, 1954.

deviations of the 1F and 2F echoes from the true bearing can be accounted for by a 1° tilt of the layer at the time.

On 5 Mc/s the two transmitting aerials compared were a $\lambda/4$ vertical aerial and a horizontal $\lambda/2$ dipole, 26 m above ground. The theoretical polar diagrams were thus quite different; the vertical aerial would have been expected to provide the stronger signals at elevations below 15°, and the horizontal aerial at all greater angles. In practice, however, very little radiation was received from either aerial at angles below about 10°, and owing to the complicated and variable nature of the echo patterns obtained on this frequency, no definite differences in the patterns could ever be related to a change of transmitting aerial. The only systematic difference noted was that the horizontal aerial gave stronger signals than the vertical one. Again, when observations were made in adjacent periods on corresponding echoes, or parts of the echo pattern, no significant differences could be found in the results obtained with the two aerials.

(5.3) Effect of Ionospheric and Magnetic Storms

A number of observations were made under conditions of ionospheric storm, and the results were examined to see whether any marked directional effects were associated with these periods; ionospheric data taken at Slough, Schwarzenburg and Lindau were used. No systematic differences in respect of either the mean bearing in a period or the spread of individual values were found between the 'storm' and 'quiet' periods. Similarly, no correlation was found between these quantities and the degree of magnetic activity as indicated by the three-hourly K figure for Abinger.

(5.4) General Statistical Analysis

Table 3 summarizes the various statistics calculated for all the results, divided into echo groups as previously described. Within each group, changes of transmitting aerial, or of ionospheric and magnetic conditions, were ignored, since these have been shown not to affect the results.

The 11 Mc/s 2F results were all obtained in day-time. The 1F high-angle, and also the 5 Mc/s, observations were taken partly during the day and partly at night (80% of the latter being taken at night), but no systematic difference was indicated,

Table 3

SUMMARY OF RESULTS

Frequency	Echo group	Number of bearings	Number of 5 min means	Number of periods	α_3	v'	v'_1	v_2	v_3	v_4	δ_3	$\Delta\delta_1$	$\Delta\delta_2$
Mc/s					deg E of N	deg^2	deg^2	deg^2	deg^2	deg^2	deg	deg	deg
11·21	1F day	6 509	717	106	139·4 ± 0·05	0·34 ± 0·01	0·095 ± 0·007	0·23 ± 0·03	0·58 ± 0·05	0·35 ± 0·05	11·1	4·9	0–17·4
11·21	1F night	1 119	126	16	139·1 ± 0·16	0·37 ± 0·03	0·22 ± 0·04	0·42 ± 0·15	0·71 ± 0·16	0·51 ± 0·16	11·7	5·8	6·0–17·1
11·21	1F HA	369	35	10	139·5 ± 0·15	0·33 ± 0·03	0·10 ± 0·03	0·24 ± 0·11	0·48 ± 0·15	0·23 ± 0·15	17·5	3·8	13·1–23·6
11·21	2F	1 304	131	32	139·3 ± 0·12	0·99 ± 0·05	0·37 ± 0·06	0·46 ± 0·11	1·44 ± 0·13	0·70 ± 0·13	23·0	4·7	17·3–26·8
5·11	I (0–½ millisec)	1 243	138	26	139·2 ± 0·14	2·15 ± 0·09	0·63 ± 0·11	0·52 ± 0·14	2·83 ± 0·17	1·19 ± 0·18	13·9	7·8	0–25·4
5·11	II (½–1 millisec)	1 256	135	31	139·1 ± 0·26	5·13 ± 0·24	2·57 ± 0·43	2·15 ± 0·55	7·25 ± 0·35	4·71 ± 0·50	21·7	6·2	8·0–31·7

Meaning of symbols:

α_3 = Grand mean bearing.
v' = Variance of individual bearings about period mean.
v'_1 = Variance of 5 min means about period mean.
v_2 = Variance of period means.
v_3 = Variance of individual bearings about grand mean.
v_4 = Variance of 5 min means about grand mean.
δ_3 = Grand mean angle of elevation.
$\Delta\delta_1$ = Mean range of angles of elevation within a period.
$\Delta\delta_2$ = Range of period-mean values of angle of elevation.

and the day and night results are therefore pooled in these cases. Similarly, the results of all seasons are included together in each group.

The most notable feature exhibited by the results is the high degree of bearing accuracy which it is possible to achieve under the conditions of these experiments. The grand mean bearing in each group is within a few tenths of a degree of the correct bearing of the transmitter (139·2°); the 5 Mc/s results are actually closest to this value, but no significance can be attached to the slightly greater error of the 11 Mc/s mean, since the different mean bearings on the two frequencies could easily be due to different lining-up or zero-setting errors of the direction-finder.

The bearings on 11 Mc/s show the smallest variations about the means. Thus, on the 1F echo in day-time the overall standard deviation of individual snap observations about the grand mean is only 0·8°, while the corresponding figures for bearings averaged over 5 min, and over an hour, are 0·6° and 0·5° respectively. The angles of elevation measured on the 5 Mc/s transmissions make it appear likely that the main echoes observed in groups I and II were respectively the 1F and 2F modes of propagation. The standard deviations for the first of these, corresponding to the 11 Mc/s day-time figures just given, are 1·7°, 1·1° and 0·7°; hence time-averaging of the results is much more efficacious in improving the accuracy in the latter case.

The 1F high-angle echo bearings are very similar to those for the low-angle ray. The 11 Mc/s 1F night results, in addition to the slightly different grand mean and the increased 5 min-mean variance compared with day-time, show that the period means themselves have a greater variance than in the day.

(5.5) Comparison with Bearing Results at Shorter Ranges

The directional variations observed at vertical incidence and at a distance of 700 km have been explained in terms of two types of ionospheric irregularity, namely large-scale tilts in the F-layer, producing the slower bearing changes, and small-scale irregularities in the lower part of the E-region, producing the rapid fluctuations. It is of interest to see how far the properties of these irregularities, as deduced from the shorter-distance experiments, are consistent with the present results. Consider first the 1F day-time bearings on 11 Mc/s; the r.m.s. value of the 'slow' bearing variation at 700 km was 0·8°, corresponding to a lateral layer tilt of 1·2° at a height of 250 km. At 2 120 km this same tilt would produce an r.m.s. bearing deviation of 0·35°, i.e. a variance of 0·12 deg^2. With the very small variances resulting from the present measurements it is not possible to make an accurate division into different components, but the value of 0·095 deg^2 makes it unlikely that the slow variance is as high as 0·12 deg^2. A value of 0·09 deg^2 would provide a better explanation of the 1F and 2F day results. For the 2F echo, geometrical analysis shows that the slow component of variance, due to uncorrelated layer tilts at the two reflection points, would be about 6·5 times its value for the 1F echo over the same path (instead of ten times as at short distances). The 1F and 2F results, for which the total within-period variances, v', are 0·34 and 0·99 deg^2 respectively, are then reasonably consistent, the residual variances (0·25 deg^2 for 1F and 0·40 deg^2 for 2F) being ascribed to the more rapid fluctuations. This rapid component of variance for the 1F echo must be attributed almost entirely to ionospheric effects, since it is known from the previous work that rapidly varying site errors are only of the order of 0·02 deg^2 at this frequency. The remaining variance is much larger than that (0·06 deg^2) expected from the vertical-incidence measurements of the rapid fluctuations, on the assumption of the simple mechanism of propagation through a single irregular layer. The extra variance can be explained qualitatively by the splitting and wave-interference effects noted with the more complex pulses

observed in the present experiments. This presumably results from the longer ionospheric paths involved, but a detailed explanation cannot be given.

With the 2F echo there may be an extra variance component due to scattering at the mid-point reflection at the earth, but from the results it does not appear to be of importance. The mid-point of the transmission path lies within a few kilometres of the coast of the Gulf of Genoa, near San Remo.

At night, the 5 min-mean variance about the period mean is greater than by day, as already mentioned. This is no doubt partly due to the fact that the ionospheric tilts change more slowly at night, so that the reduction of variance achieved by time-averaging over 5 min is smaller than in the day-time; but, in addition, it appears that the lateral deviation component of variance itself must be greater at night. The results are best explained by taking this component to be about $0{\cdot}16\,\text{deg}^2$, although a value of $0{\cdot}12\,\text{deg}^2$ would not be inconsistent with the results. No evidence for this night-time increase was found in the previous experiments at shorter distances, which indicated that, if anything, the tilts decreased slightly at night. The present results also indicate a slight decrease of the rapid variance component at night, as shown by the larger day-time value of $(v' - v'_1)$, but this difference is not statistically significant.

It is possible to account for the 5 Mc/s night-time results in terms of a slow component of variance which is the same as on 11 Mc/s, plus a much larger rapid component, namely about $2\,\text{deg}^2$ for group I and $4\,\text{deg}^2$ for group II. These large values are clearly caused by the mutual interference of the components present in the strobed part of the echo pattern.

(5.6) Comparison with Observations on a Narrow-Aperture Direction-Finder, using Pulse and C.W. Transmissions

During the experiments a number of bearing observations were made over the same periods of time on a standard type of 4-aerial Adcock direction-finder with cathode-ray direction-finder display, and a comparison of the results shows the marked superiority of the wide-aperture system. For eight periods on the 11 Mc/s 1F echo in day-time, the values of v' were $1{\cdot}6 \pm 0{\cdot}1\,\text{deg}^2$ and $0{\cdot}17 \pm 0{\cdot}02\,\text{deg}^2$ on the Adcock and the spaced-loop system respectively. Similarly, for three night periods on 5 Mc/s (group I) the corresponding results were $18{\cdot}2 \pm 2{\cdot}3\,\text{deg}^2$ and $1{\cdot}4 \pm 0{\cdot}2\,\text{deg}^2$.

A few periods of c.w. transmission were also provided during the experiments, and observations were again made on both direction-finders. Only the rapid bearing fluctuations were studied, the procedure being to take readings every 5 sec over periods of 5 min. The results of a number of such periods were then pooled, and the overall variance about the period mean was calculated as before. On 11 Mc/s, 15 day-time periods were used, with pulse patterns varying from a simple 1F and 2F structure to those shown in Figs. 1(*a*) and 1(*b*). The resulting variances were $8{\cdot}9 \pm 0{\cdot}6\,\text{deg}^2$ for the Adcock direction-finder and $2{\cdot}1 \pm 0{\cdot}2\,\text{deg}^2$ for the spaced-loop direction-finder. On 5 Mc/s, three night periods were used, the echo patterns being similar in character to the second example shown in Fig. 1(*c*), but extending over only 1–1½ millisec. The variance obtained with the Adcock direction-finder was now $25{\cdot}0 \pm 2{\cdot}9\,\text{deg}^2$, and that for the spaced-loop direction-finder was $16{\cdot}9 \pm 2{\cdot}6\,\text{deg}^2$.

Thus, in all cases the wide-aperture spaced-loop system gave much smaller variances than the Adcock system. It is also seen that a substantial variance reduction was achieved with either equipment by using a selected portion of a pulse pattern instead of a c.w. transmission.

(6) CONCLUSIONS

The observations have shown that the bearing errors due to ionospheric influences on a transmission path of 2 120 km are extremely small in general. The grand mean bearing on 11 Mc/s was within $0{\cdot}2°$, and on 5 Mc/s within $0{\cdot}1°$, of the correct value. The bearing deviations can be divided into slowly and rapidly varying components, as at shorter distances. The slow component is approximately of the magnitude expected from the previous measurements of layer tilts, and for first-order F-reflection is smaller than the rapid component. The total standard deviation of snap observations about an hourly mean for this echo on 11 Mc/s is only $0{\cdot}6°$; the standard deviation of the hourly means themselves has about the same value, and a seasonal variation contributes largely to this.

On 5 Mc/s, at night, simultaneous E- and F-layer reflections produce complex echo patterns, resulting in considerably larger bearing fluctuations. These are principally of the rapid kind, owing to overlapping of the components.

The results are unaffected by magnetic or ionospheric storms, or by such changes in transmitting aerial as have been tried. The bearing variances are in all cases appreciably less than those obtained with a narrow-aperture Adcock direction-finder, or when using c.w. transmissions.

(7) ACKNOWLEDGMENTS

The author acknowledges the assistance of a number of colleagues, particularly Messrs. A. Ambery and L. A. Bonvini, in making the observations. Mr. Bonvini also carried out much of the reduction and numerical analysis of the results.

The work described above was carried out as part of the programme of the Radio Research Board. The paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research.

(8) REFERENCES

(1) BRAMLEY, E. N.: 'Direction Finding Studies of Large-Scale Ionospheric Irregularities', *Proceedings of the Royal Society* A, 1953, **220**, p. 39.

(2) BRAMLEY, E. N.: 'Some Aspects of the Rapid Directional Fluctuations of Short Radio Waves Reflected at the Ionosphere', *Proceedings I.E.E.*, Paper No. 1713 R, August, 1954 (**102**, B, p. 533).

(3) ROSS, W., BRAMLEY, E. N., and ASHWELL, G. E.: 'A Phase Comparison Method of Measuring the Direction of Arrival of Ionospheric Radio Waves', *ibid.*, Paper No. 1134, July, 1951 (**98**, Part III, p. 294).

(4) BRAMLEY, E. N., and ROSS, W.: 'Measurements of the Direction of Arrival of Short Radio Waves Reflected from the Ionosphere', *Proceedings of the Royal Society* A, 1951, **207**, p. 251.

(5) TIPPETT, L. H. C.: 'On the Extreme Individuals and the Range of Samples taken from a Normal Population', *Biometrika*, 1925, **17**, p. 364.

621.372.8.017

Paper No. 1969 R
May 1956

A NEW TREATMENT OF LOSSY PERIODIC WAVEGUIDES

By P. N. BUTCHER, Ph.D.

(*The paper was first received 26th July, and in revised form 21st October*, 1955.)

SUMMARY

The conventional treatment of propagation coefficients in lossy periodic waveguides suffers from certain major defects. It gives no information about the effects of the losses on the phase-change coefficient, it breaks down when the frequency approaches the edge of a pass-band from within and it does not work at all in a stop-band. A new treatment is described which removes all these defects. It is based on the following result: the propagation coefficient of a mode in a lossy guide at the frequency ω is equal to the propagation coefficient of the corresponding mode in the lossless guide at the frequency $\omega(1 - j/2Q_c)$. Here, Q_c is the "complex Q-factor" of the mode at the frequency ω. It is given by an explicit formula which holds good at all frequencies when the losses are small. When ω lies within a pass-band Q_c is equal to ω times the mean energy stored in a period of the guide divided by the complex power dissipated in the same period. When ω lies in a stop-band Q_c is equal to the analytic continuation of its values in the pass-bands.

LIST OF PRINCIPAL SYMBOLS

a_s = Coefficients.
α_l = Attenuation coefficient in the lossy guide.
β_l = Phase-change coefficient in the lossy guide.
D = Period of the guide.
D_{rs} = (r, s)th matrix element in the equations for a_s.
δ = Skin-depth of the metal.
$E(\gamma)$ = Electric vector in the lossless guide (suffix added when the mode is degenerate).
$E_l(\gamma_l)$ = Electric vector in the lossy guide.
γ = Propagation coefficient in the lossless guide at the frequency ω.
γ_0 = Propagation coefficient in the lossless guide at which the group velocity vanishes.
Γ = Propagation coefficient in the lossless guide at the frequency ω_l.
γ_l = Propagation coefficient in the lossy guide at the frequency ω_l.
$H(\gamma)$ = Magnetic vector in the lossless guide (suffix added when the mode is degenerate).
$Hl(\gamma_l)$ = Magnetic vector in the lossy guide.
K = Positive constant.
$\epsilon_m, \mu_m, \sigma_m$ = Permittivity, permeability and conductivity of the metal.
$\epsilon_d, \mu_d, \sigma_d$ = Permittivity, permeability and conductivity of the dielectric.
l_0 = Attenuation length.
m = Number of half-wavelengths in a short-circuited symmetrical structure.
N = Number of periods in a short-circuited symmetrical structure.
P_0 = Mean power-flow across the plane $z = 0$.
P_l = Mean power dissipated ohmically in a period of the guide.
Q, Q_w = Real Q-factors (total and wall).
Q_c, Q_{cd}, Q_{cw} = Complex Q-factors (total, dielectric and wall).
R = Degeneracy.
t = Time.
v_g = Group velocity.
W = Mean energy stored in a period of the guide.
$\omega, \omega(\gamma)$ = Frequency in the lossless guide when the propagation coefficient is γ.
$\omega_l, \omega_l(\gamma_l)$ = Frequency in the lossy guide when the propagation coefficient is γ_l.
ω_0 = Frequency at which the group velocity vanishes.
x, y, z = Cartesian co-ordinates.

(1) INTRODUCTION

The most important characteristic of a mode in a periodic waveguide is its propagation coefficient, which is a function of the frequency. The theoretical determination of this function is made much easier by supposing that the guide is lossless. For many purposes, however, the calculation must be refined so as to take account of the losses. The familiar treatment of this problem which is given in many textbooks (e.g. Reference 1) is severely limited in scope, since it gives no information about the effects of the losses on the phase-change coefficient, it breaks down when the frequency approaches the edge of pass-band from within and it does not work at all in a stop-band. This paper describes a new treatment which removes all these defects. It will be particularly useful in the design of electronic devices operating near the edge of a pass-band, the π-mode linear accelerator for example.

(2) THE PROBLEM

Consider a periodic waveguide with period D which is constructed of a highly conducting metal with permittivity ϵ_m, permeability μ_m and conductivity σ_m, and is filled with a slightly lossy dielectric with permittivity ϵ_d, permeability μ_d and conductivity σ_d. Let (x, y, z) be the Cartesian co-ordinates of a point in space, let the z-axis be parallel to the axis of the guide and let the origin be located at some geometrically convenient point in the guide. It will be convenient to work with complex field vectors which are functions of (x, y, z) only. It will be understood that the real field vectors are obtained by taking the real parts of the complex vectors times $\exp(j\omega t)$, where ω is the angular frequency of the field considered and t is the time.

In the particular case of an ordinary uniform guide, the electromagnetic behaviour of the structure is most easily discussed in terms of its "exponential" modes of oscillation. These depend on z through the factor $\exp(-\gamma z)$, where γ is the propagation coefficient. This type of mode is characteristic of uniform guides and cannot be supported by a non-uniform periodic structure. In the latter case, however, we can make an obvious generalization and introduce "quasi-exponential" modes which depend on z in such a way that the field vectors at $(x, y, z + D)$ differ from those at (x, y, z) only by a constant factor $\exp(-\gamma D)$. We still call γ the propagation coefficient; its real part is the attenuation coefficient and its imaginary part is the phase-change coefficient. For each such mode the propagation coefficient is a function of the frequency. Because the effects of the ohmic losses in the metal and the dielectric are usually small it is

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Dr. Butcher is at the Radar Research Establishment.

sufficient to determine this function by first of all neglecting the losses entirely and then taking them into account by means of a suitable approximation. The paper is concerned with the second step, the first will be assumed to have been accomplished.

Some care must be taken with the notation in order to simplify the mathematical analysis of this problem. We shall reserve ω and γ for the frequency and propagation coefficient of a mode in the lossless guide and add a suffix l when referring to the corresponding mode in the lossy guide. It will be convenient to suppose that the real part of ω_l is positive. This can be done without any loss of generality. We are, of course, most interested in the case in which the imaginary part of ω_l is zero. However, this restriction will only be imposed in the new treatment when it is necessary to point out some special property of this case, and in the conventional treatment in order to shorten what is probably a familiar argument. In both the lossless and the lossy guide the frequency is a function of the propagation coefficient. When it is necessary to show this explicitly we shall write $\omega(\gamma)$ and $\omega_l(\gamma_l)$ for the lossless and the lossy guide respectively. The attenuation coefficient and phase-change coefficient in the lossy guide will be denoted by α_l and β_l respectively, the frequency to which they relate being made clear in the context. Finally, in the course of the calculation of the propagation coefficient in the *lossy* guide at the frequency ω_l it will be necessary to refer to the propagation coefficient in the *lossless* guide when ω is equal to ω_l. This will be denoted by Γ.

(3) THE CONVENTIONAL TREATMENT

Before giving the new approach to the calculation of the propagation coefficient in a lossy guide, we must briefly consider the conventional treatment. Several authors have given this for uniform guides (e.g. Reference 1). The extension to non-uniform periodic guides is almost trivial, indeed it has often been tacitly assumed, but it does not appear to be given in the literature. The basis of the treatment is the conservation of energy in the section of the lossy guide between the planes $z = 0$ and $z = D$. When ω_l is real, the mean energy stored in this section is constant in time. Because the mode is quasi-exponential, the mean power flow across the plane $z = D$ is $\exp(-2\alpha_l D)$ times the mean power flow P_0 across the plane $z = 0$. The net mean power input to the section through these two planes is equal to the mean power P_l dissipated ohmically in the dielectric and the metal. Hence

$$P_0[1 - \exp(-2\alpha_l D)] = P_l \quad . \quad . \quad . \quad . \quad (1)$$

If the attenuation coefficient is small enough, we can expand the exponential as a power series in $2\alpha_l D$ and neglect powers higher than the first. This approximation is usually valid and leads to the equation

$$\alpha_l = P_l/2P_0D \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

It is not difficult to approximate to P_l in terms of the field in the lossless guide and, again, the approximation is usually valid. However, when we try to approximate to P_0 the situation is much less satisfactory. If ω_l is well within a pass-band P_0 can be taken to be the mean longitudinal power flow in the lossless guide. It can be shown that this is equal to the group velocity v_g times the mean energy stored per unit length W/D, where W is the mean energy stored in a period D. Hence eqn. (2) becomes

$$\alpha_l = P_l/2Wv_g = \omega_l/2Qv_g \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where

$$Q = \omega_l W/P_l \quad . \quad . \quad . \quad . \quad . \quad (4)$$

is the "real" Q-factor of the mode considered at the frequency ω_l. The adjective "real" is used here in order to distinguish Q from the "complex" Q-factor which will be introduced later on. As ω_l approaches the edge of a pass-band from within, P_0 tends to zero in the lossless guide but remains finite in the lossy guide. Hence the above approximation becomes worse and worse leading finally to the physically unacceptable result that α_l is infinite at the edge of a pass-band. If ω_l lies within a stop-band P_0 is precisely zero in the lossless guide (but not in the lossy guide) and the above approximation cannot even be attempted. Finally, we note that even when the conventional treatment is valid it does not provide any information about the effects of the losses on the phase-change coefficient.

(4) THE NEW TREATMENT

We want to find γ_l when ω_l is given. Now, in the lossless guide γ is a known function of ω, hence, if we can find ω such that γ is equal to γ_l, our problem is solved. This can be achieved after a little manipulation of Maxwell's equations, as shown below.

Let $\boldsymbol{E}_l(\gamma_l)$ and $\boldsymbol{H}_l(\gamma_l)$ be the complex vectors of a mode in the lossy guide with propagation coefficient γ_l and frequency $\omega_l(\gamma_l)$. In the lossy dielectric these vectors satisfy Maxwell's equations:

$$\text{curl } \boldsymbol{E}_l(\gamma_l) + j\omega_l(\gamma_l)\mu_d \boldsymbol{H}_l(\gamma_l) = 0 \quad . \quad . \quad . \quad (5)$$

$$\text{curl } \boldsymbol{H}_l(\gamma_l) - j\omega_l(\gamma_l)\epsilon_d \boldsymbol{E}_l(\gamma_l) = \sigma_d \boldsymbol{E}_l(\gamma_l) \quad . \quad . \quad (6)$$

Let $\boldsymbol{E}(-\gamma_l)$ and $\boldsymbol{H}(-\gamma_l)$ be the complex vectors of the corresponding mode in the lossless guide with propagation coefficient γ equal to $-\gamma_l$. The reason for using this "reversed wave" will become clear in a moment. It has the same frequency, namely $\omega(\gamma_l)$, as the "direct wave" with propagation coefficient $+\gamma_l$ (see Section 8.1). Hence, in the lossless dielectric we have Maxwell's equations

$$\text{curl } \boldsymbol{E}(-\gamma_l) + j\omega(\gamma_l)\mu_d \boldsymbol{H}(-\gamma_l) = 0 \quad . \quad . \quad (7)$$

$$\text{curl } \boldsymbol{H}(-\gamma_l) - j\omega(\gamma_l)\epsilon_d \boldsymbol{E}(-\gamma_l) = 0 \quad . \quad . \quad . \quad (8)$$

Taking the scalar product of eqns. (5), (6), (7) and (8) with $-\boldsymbol{H}(-\gamma_l)$, $-\boldsymbol{E}(-\gamma_l)$, $\boldsymbol{H}_l(\gamma_l)$ and $\boldsymbol{E}_l(\gamma_l)$ respectively, adding and making use of the well-known formula for the divergence of a vector product we obtain

$$\text{div}\,[\boldsymbol{E}(-\gamma_l)\times\boldsymbol{H}_l(\gamma_l) - \boldsymbol{E}_l(\gamma_l)\times\boldsymbol{H}(-\gamma_l)] + \sigma_d \boldsymbol{E}(-\gamma_l)\,.\,\boldsymbol{E}_l(\gamma_l)$$
$$= j[\omega(\gamma_l) - \omega_l(\gamma_l)][\epsilon_d \boldsymbol{E}(-\gamma_l)\,.\,\boldsymbol{E}(\gamma_l) - \mu_d \boldsymbol{H}(-\gamma_l)\,.\,\boldsymbol{H}_l(\gamma_l)] \quad . \quad . \quad . \quad . \quad . \quad (9)$$

This equation can be integrated over the region V of the dielectric bounded by the planes $z = 0$ and $z = D$ and the metal surfaces S. The volume integral of the divergence can be converted into a surface integral taken over the boundary. It is easy to see that the contributions to the surface integral from the planes $z = 0$ and $z = D$ cancel out because the two fields involved are quasi-exponential modes with oppositely signed propagation coefficients. This cancellation was arranged, by taking the reversed wave in the lossless guide, because the contribution from these two planes is otherwise difficult to approximate. There remains the integral over the metal surfaces to which only the second term in the divergence contributes because $\boldsymbol{E}(-\gamma_l)$ is normal to S. Thus, after rearranging, we obtain the desired formula for the frequency in the lossless guide at which γ is equal to γ_l:

$$\omega(\gamma_l) = \omega_l(\gamma_l)(1 - j/2Q_c) \quad . \quad . \quad . \quad . \quad (10)$$

where

$$\frac{1}{Q_c} = \frac{-\int_S \boldsymbol{E}_l(\gamma_l)\times\boldsymbol{H}(-\gamma_l)\,.\,\boldsymbol{n}dS + \int_V \sigma_d \boldsymbol{E}(-\gamma_l)\,.\,\boldsymbol{E}_l(\gamma_l)dV}{\frac{\omega_l(\gamma_l)}{2}\int_V [\epsilon_d \boldsymbol{E}(-\gamma_l)\,.\,\boldsymbol{E}_l(\gamma_l) - \mu_d \boldsymbol{H}(-\gamma_l)\,.\,\boldsymbol{H}_l(\gamma_l)]dV} \quad . \quad . \quad . \quad . \quad . \quad (11)$$

in which $\boldsymbol{n}$ is a unit vector, normal to S and directed into the metal. Eqns. (10) and (11) are exact. An approximation to Q_c is derived in Section 5.

(5) THE COMPLEX Q-FACTOR

Suppose that ω_l is given. Now, in eqn. (11), γ_l is the propagation coefficient in the lossy guide at the frequency ω_l and is not yet known. However, if the losses are small, γ_l is very close to the propagation coefficient Γ in the *lossless* guide at the given frequency ω_l. Using Γ in eqn. (11) in place of γ_l will only result in a second-order error because both σ_d and the components of $\boldsymbol{E}_l(\gamma_l)$ tangential to S are small. Then, except in the case of $\boldsymbol{E}_l(\Gamma)$ in the surface integral, which is dealt with later, we can replace the field vectors in the lossy guide by those in the lossless guide and still the error in eqn. (11) will be only of the second order. Thus we obtain

$$\frac{1}{Q_c} = \frac{-\int_S \boldsymbol{E}_l(\Gamma)\times\boldsymbol{H}(-\Gamma)\,.\,\boldsymbol{n}dS + \int_V \sigma_d \boldsymbol{E}(-\Gamma)\,.\,\boldsymbol{E}(\Gamma)dV}{\frac{\omega_l}{2}\int_V [\epsilon_d \boldsymbol{E}(-\Gamma)\,.\,\boldsymbol{E}(\Gamma) - \mu_d \boldsymbol{H}(-\Gamma)\,.\,\boldsymbol{H}(\Gamma)]dV} \quad . \quad . \quad . \quad . \quad . \quad (12)$$

where Γ is the propagation coefficient in the lossless guide at the given frequency ω_l, which can be calculated if the relation between frequency and propagation coefficient in the lossless guide is known.

When ω_l is real and lies within a pass-band (through the lossless guide) of the mode considered, Q_c has a simple interpretation. In this case Γ is imaginary and

$$\boldsymbol{E}(-\Gamma) = \boldsymbol{E}^*(\Gamma)$$
$$\boldsymbol{H}(-\Gamma) = -\boldsymbol{H}^*(\Gamma) \quad . \quad . \quad . \quad . \quad (13)$$

(see Section 8.1). Making this substitution in eqn. (12) we obtain

$$\frac{1}{Q_c} = \frac{\int_S \boldsymbol{E}_l(\Gamma)\times\boldsymbol{H}^*(\Gamma)\,.\,\boldsymbol{n}dS + \int_V \sigma_d |\boldsymbol{E}(\Gamma)|^2 dV}{\frac{\omega_l}{2}\int_V [\epsilon_d |\boldsymbol{E}(\Gamma)|^2 + \mu_d |\boldsymbol{H}(\Gamma)|^2]dV}$$

$$= \left(\frac{1}{\omega_l}\right)\frac{\text{complex power dissipated in the period}}{\text{mean energy stored in the period}} \quad (14)$$

to the present degree of approximation. Thus, the real part of the reciprocal of Q_c is just the reciprocal of the real Q-factor defined by eqn. (4). In view of this result it is appropriate to call Q_c the "complex Q-factor" of the mode considered at the frequency ω_l. It should be emphasized, however, that eqn. (14) is valid only when ω_l is real and lies within a pass-band, while eqn. (12) holds good for all values of ω_l. In other words, eqn. (12) is the analytic continuation of eqn. (14).

We obtain the final approximation to Q_c by using the familiar formula for the components of $\boldsymbol{E}_l(\Gamma)$ tangential to the metal surface:[1]

$$\boldsymbol{E}_{lt}(\Gamma) = \boldsymbol{H}(\Gamma)\times\boldsymbol{n}(1+j)/\sigma_m\delta \quad . \quad . \quad . \quad (15)$$

where $\delta = \sqrt{(2/\sigma_m\mu_m\omega_l)}$ is the skin-depth of the metal at the frequency ω_l, which, we note, will be complex if ω_l is complex. Thus eqn. (12) becomes

$$\frac{1}{Q_c} = \frac{1}{Q_{cw}} + \frac{1}{Q_{cd}} \quad . \quad . \quad . \quad . \quad . \quad (16)$$

where, since $\boldsymbol{H}(-\Gamma)$ is tangential to S,

$$\frac{1}{Q_{cw}} = \frac{-(1+j)\int_S (\sigma_m\delta)^{-1}\boldsymbol{H}(-\Gamma)\,.\,\boldsymbol{H}(\Gamma)dS}{\frac{\omega_l}{2}\int_V [\epsilon_d \boldsymbol{E}(-\Gamma)\,.\,\boldsymbol{E}(\Gamma) - \mu_d \boldsymbol{H}(-\Gamma)\,.\,\boldsymbol{H}(\Gamma)]dV} \quad (17)$$

and

$$\frac{1}{Q_{cd}} = \frac{\int_V \sigma_d \boldsymbol{E}(-\Gamma)\,.\,\boldsymbol{E}(\Gamma)dV}{\frac{\omega_l}{2}\int_V [\epsilon_d \boldsymbol{E}(-\Gamma)\,.\,\boldsymbol{E}(\Gamma) - \mu_d \boldsymbol{H}(-\Gamma)\,.\,\boldsymbol{H}(\Gamma)]dV} \quad (18)$$

Q_{cw} and Q_{cd} are the complex Q-factors when the losses arise entirely in the walls and entirely in the dielectric respectively. They can be calculated if the fields in the lossless guide are known.

There are several things to notice about Q_{cw} and Q_{cd}. We may write

$$\frac{1}{Q_{cw}} = \frac{1+j}{Q_w} \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

When ω_l is real and lies within a pass-band it is obvious from eqns. (13), (17), (18) and (19) that Q_w and Q_{cd} are both real and positive. Moreover, it is clear from our previous remarks that they are respectively equal to the real Q-factors given by eqn. (4) when the losses arise entirely in the walls and entirely in the dielectric respectively. It may be shown that they remain real when ω_l lies in a stop-band, but they no longer have this simple interpretation. The denominators in eqns (17) and (18) can be simplified slightly by using the identity

$$\int_V \epsilon_d \boldsymbol{E}(-\Gamma)\,.\,\boldsymbol{E}(\Gamma)dV = -\int_V \mu_d \boldsymbol{H}(-\Gamma)\,.\,\boldsymbol{H}(\Gamma)\,dV \quad (20)$$

which is derived in Section 8.1. This is a generalization of the well-known equality between the mean stored electric and magnetic energies to which it reduces when ω_l is real and lies within a pass-band.

All the preceding analysis is valid when the electromagnetic constants of both the metal and the dielectric are functions of position. The case in which they are independent of position is, however, particularly important in practice. Then, using eqn. (20), we find that eqns. (17) and (18) reduce to

$$\frac{1}{Q_{cw}} = (1+j)\frac{\delta\mu_m}{2\mu_d}\frac{\int_S \boldsymbol{H}(-\Gamma)\,.\,\boldsymbol{H}(\Gamma)dS}{\int_V \boldsymbol{H}(-\Gamma)\,.\,\boldsymbol{H}(\Gamma)dV} \quad . \quad . \quad (21)$$

and

$$\frac{1}{Q_{cd}} = \frac{\sigma}{\omega_l\epsilon_d} \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

Hence, when ω_l is given a real value, which may lie in a pass-band or in a stop-band, Q_{cd} reduces to the real Q-factor of the dielectric as usually defined.[3]

(6) DISCUSSION

(6.1) The Linear Approximation

The propagation coefficient in the lossy guide at the frequency ω_l is equal to that in the lossless guide at the frequency $\omega(\gamma_l)$

given by eqn. (10). If the losses are small, Q_c is large and $\omega(\gamma_l)$ is very close to ω_l. Hence, from Taylor's theorem we have approximately

$$\gamma_l = \Gamma - \frac{j\omega_l}{2Q_c}\frac{d\gamma}{d\omega}\bigg|_{\omega_l} \quad . \quad . \quad . \quad . \quad . \quad (23)$$

We shall call this the linear approximation. It can be written in the more perspicuous form

$$\gamma_l = \Gamma + \frac{\omega_l}{2v_g}\frac{1}{Q_c} \quad . \quad . \quad . \quad . \quad . \quad (24)$$

where

$$v_g = j\Big/\frac{d\gamma}{d\omega}\bigg|_{\omega_l} = j\frac{d\omega}{d\gamma}\bigg|_{\Gamma} \quad . \quad . \quad . \quad . \quad . \quad (25)$$

If ω_l is real and lies within a pass-band, Γ is imaginary and v_g is just the group velocity. In this case the real part of eqn. (24) gives for the attenuation coefficient in the lossy guide

$$\alpha_l = \frac{\omega_l}{2v_g}\mathscr{R}\left(\frac{1}{Q_c}\right) \quad . \quad . \quad . \quad . \quad . \quad (26)$$

This is the same as eqn. (3), which was obtained by the conventional treatment. However, the imaginary part of eqn. (24) gives the corrected phase-change coefficient in the lossy guide:

$$\beta_l = \mathscr{I}(\Gamma) + \frac{\omega_l}{2v_g}\mathscr{I}\left(\frac{1}{Q_c}\right) \quad . \quad . \quad . \quad . \quad (27)$$

When the losses are entirely in the walls, the real and imaginary parts of $1/Q_c$ are both equal to $1/Q_w$ [eqn. (19)]. Hence the increment of the phase-change coefficient is equal to the attenuation coefficient in the lossy guide. When dielectric loss is also present, the attenuation coefficient is naturally increased but the increment of the phase-change coefficient is not affected because Q_{cd} is real. It is easy to show that the phase-velocity is reduced in magnitude by the ohmic losses when the phase and group velocities in the lossless guide have the same direction. This is the case most often met in practice. Sometimes, however, the phase and group velocities in the lossless guide are oppositely directed. Then the phase-velocity is increased in magnitude by the ohmic losses.

If ω_l is real and lies within a stop-band, the imaginary part of Γ is fixed at either 0 or π/D and v_g is imaginary. In this case the real and imaginary parts of eqn. (24) give

$$\left.\begin{aligned}\alpha_l &= \mathscr{R}(\Gamma) - \frac{\omega_l}{2jv_g}\mathscr{I}\left(\frac{1}{Q_c}\right)\\ \beta_l &= \mathscr{I}(\Gamma) + \frac{\omega_l}{2jv_g}\mathscr{R}\left(\frac{1}{Q_c}\right)\end{aligned}\right\} \quad . \quad . \quad . \quad (28)$$

When the losses are entirely in the walls, the corrections to the attenuation and phase-change coefficients are equal in magnitude but opposite in sign. When dielectric loss is also present, the correction to the phase-change coefficient is increased but the correction to the attenuation coefficient is not affected.

(6.2) Failure of the Linear Approximation

When v_g tends to zero, as it does at the edge of a pass-band, the linear approximation, eqn. (24), suggests that the increment of the propagation coefficient resulting from the ohmic losses tends to infinity. In fact, of course, the cut-off frequencies at which v_g is zero are just those for which $d\gamma/d\omega$ is infinite and the linear approximation becomes increasingly inaccurate when ω_l approaches such a value. To consider this matter further, suppose that, in the neighbourhood of a real cut-off frequency ω_0, the relation between γ and ω is given with sufficient accuracy by the equation

$$\omega - \omega_0 = K(\gamma - \gamma_0)^2 \quad . \quad . \quad . \quad . \quad (29)$$

Here K is a real constant which, to be definite, we shall take to be positive. Using this equation amounts to fitting a parabola to the dispersion curve of the lossless guide near the frequency ω_0 at which it has a horizontal tangent. This is often a good approximation. Putting ω equal to $\omega_l(1 - j/2Q_c)$ and solving for γ we obtain the propagation constant in the lossy guide at the frequency ω_l:

$$\gamma_l = \gamma_0 + [(\omega_l - \omega_0 - j\omega_l/2Q_c)/K]^{1/2} \quad . \quad . \quad (30)$$

If $|\omega_l - \omega_0|$ is much greater than $|\omega_l/2Q_c|$ we can expand the square root in eqn. (30) as a power series in $j\omega_l/2Q_c$, neglect powers higher than the first and reproduce the linear approximation of eqn. (24). If this inequality is not valid, neither is the expansion and γ_l must be left in the more complicated form of eqn. (30). A particularly interesting case is that in which the losses arise only in the walls and ω_l is set equal to ω_0. Then we find from eqns. (19) and (30) that

$$\gamma_l = \gamma_0 + \left(\frac{\omega_0}{KQ_w\sqrt{2}}\right)^{1/2}\exp(-j\pi/8) \quad . \quad . \quad (31)$$

When ω_l is real, the above condition for the validity of the linear approximation can easily be stated in another way which is worth noting. Consider the mode in the *lossy* guide with propagation coefficient Γ (i.e. with γ_l equal to the propagation constant in the *lossless* guide at the *given* frequency ω_l) and frequency $\omega_l(\Gamma)$. We can find this frequency from eqn. (10) by putting γ_l equal to Γ. Then $\omega(\gamma_l)$ reduces to ω_l and so

$$\begin{aligned}\omega_l(\Gamma) &= \omega_l/(1 - j/2Q_c)\\ &\simeq \omega_l(1 + j/2Q_c) \quad . \quad . \quad . \quad . \quad (32)\end{aligned}$$

if Q_c is large. In this case $\omega_l(\Gamma)$ is very close to the given frequency ω_l and we can use the approximation to Q_c which was obtained in Section 5. It is clear from eqn. (32) that the mode has a time-constant equal to the reciprocal of $\mathscr{R}[\omega_l/2Q_c]$. Consequently it will have spectral bandwidth between the half-power points equal to twice the reciprocal of the time-constant, i.e. $\mathscr{R}[\omega_l/Q_c]$. This has the same order of magnitude as $|\omega_l/Q_c|$. Hence, the linear approximation will be valid provided the separation between ω_l and cut-off is much greater than the bandwidth of the mode in the lossy guide with the propagation constant Γ.

(6.3) Short-Circuited Symmetrical Structures

Structures which are unchanged by reflection in the plane $z = 0$ often occur in practice. We shall call them symmetrical structures. Consider such a structure. The transverse (x and y) components of the electric field in the direct wave, with propagation coefficient γ_l, at the point (x, y, z) are identical with those in the reversed wave, with propagation coefficient $-\gamma_l$, at the point $(x, y, -z)$. This is obvious from an inspection of Maxwell's equations, provided we scale the two waves appropriately. It follows that the standing-wave system formed by the difference of the direct and reversed waves will have zero transverse electric field on the plane $z = 0$. If γ_l is equal to $jm\pi/ND$, where m and N are positive integers, then both the direct and reversed waves will advance in phase by $m\pi$ on going from $z = 0$ to $z = ND$ and the standing-wave system will also have zero transverse electric field on the plane $z = ND$. The insertion of perfectly conducting plates in these two planes will, consequently, disturb neither the field distribution nor the frequency. Thus we obtain the so-

called $m\pi/N$ mode of a short-circuited symmetrical structure N periods in length. If we neglect the losses, its frequency is just $\omega(jm\pi/ND)$. If we take the losses into account, its frequency is $\omega_l(jm\pi/ND)$, which can be obtained from eqn. (10) by putting γ_l equal to $jm\pi/ND$ and taking Q_c from Section 5 with ω_l set equal to $\omega(jm\pi/ND)$. The phase-change coefficient can be restricted to lie between 0 and π/D for the purpose of this discussion. Hence it is necessary to consider only values of m from 0 to N inclusive. In general there is a cavity mode corresponding to each of these values. The values 0 and N require special consideration, however, and may sometimes be excluded.

Slater has defined the attenuation length of a mode in a lossy guide as

$$l_0 = 1/2\alpha_l \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

(Reference 2, p. 485). However, he uses the approximate formula (3) to calculate α_l. This breaks down at the cut-off frequencies because v_g is zero there. Slater was particularly concerned with the cut-off frequency of a symmetrical structure at which the propagation coefficient is $j\pi/D$ in the lossless guide. *Ad hoc*, he defined the attenuation length at this frequency as that length of the lossless guide which, when short-circuited by perfectly conducting end-plates, makes the frequency difference between the π-mode resonance and its nearest neighbour comparable to the frequency breadth of the π-mode resonance when the guide is lossy (Reference 2, p. 496). Let the short-circuited section be N periods in length. We shall use eqn. (29) with ω_0 set equal to the frequency of the π-mode resonance to represent the relation between the frequency and propagation coefficient in the lossless guide near the band-edge. The frequency of the nearest neighbour to the π-mode is obtainable by putting γ equal to $j(N-1)\pi/ND$ in eqn. (29). Clearly it differs from ω_0 by $K(\pi/ND)^2$. If the losses are entirely due to the walls the bandwidth of the π-mode resonance is ω_0/Q_w. Equating this to $K(\pi/ND)^2$ and solving for the attenuation length ND we find that the above definition leads to the value

$$l_0 = \pi\left(\frac{KQ_w}{\omega_0}\right)^{1/2} \quad . \quad . \quad . \quad . \quad . \quad (34)$$

when we interpret "comparable to" as meaning "equal to." This is larger than the value obtained from eqns. (31) and (33) by the factor $2^{\frac{1}{2}}\pi \cos \pi/8$, which is equal to 4·88.

(6.4) Degenerate Modes

Throughout our discussion it has been tacitly assumed that the field of the mode in the lossy guide is only slightly different from that of the corresponding mode in the lossless guide. This will obviously be so when the losses are small provided that there is no other mode in the lossless guide with the same frequency and propagation coefficient. This is the most important case in practice. However, sometimes there is more than one mode in the lossless guide with the same frequency and propagation coefficient, i.e. the modes are degenerate. These modes will then give rise to an equal number of modes in the lossy guide. At any given frequency the latter have, in general, slightly different propagation coefficients and fields which are slightly different from certain linear combinations of the degenerate modes in the lossless guide. These linear combinations are the "corresponding" modes in the lossless guide which must be used in the evaluation of the complex Q-factors of the modes in the lossy guide. To find them we must make a rather more complicated application of the method of Sections 4 and 5. This is given in Section 8.3. There, a determinantal equation for the complex Q-factors is derived which enables them to be determined without a prior calculation of the appropriate linear combinations of the modes in the lossless guide.

(7) REFERENCES

(1) SLATER, J. C.: "Microwave Electronics" (D. Van Nostrand Co. Inc., New York, 1954), Section 3.4.

(2) SLATER, J. C.: "The Design of Linear Accelerators," *Reviews of Modern Physics*, 1948, **20**, p. 473.

(3) SCHELKUNOFF, S. A.: "Electromagnetic Waves" (D. Van Nostrand Co. Inc., New York, 1943), p. 83.

(4) PAPADOPOULOS, V. M.: "Propagation of Electromagnetic Waves in Cylindrical Wave-Guides with Imperfectly Conducting Walls," *Quarterly Journal of Mechanics and Applied Mathematics*, 1954, **7**, p. 326.

(8) APPENDICES

(8.1) Some Properties of the Modes in a Lossless Guide

In this Appendix some of the properties of the modes in a lossless guide which were assumed in the text will be verified. First of all we shall derive a simple relation between the complex vectors of the direct and reversed waves which is valid when the propagation constant is imaginary. Let $\boldsymbol{E}(\gamma)$ and $\boldsymbol{H}(\gamma)$ be the complex vectors of a mode in the lossless guide with propagation coefficient γ and frequency $\omega(\gamma)$. In the lossless dielectric these vectors satisfy Maxwell's equations

$$\text{curl } \boldsymbol{E}(\gamma) + j\omega(\gamma)\mu_d \boldsymbol{H}(\gamma) = 0 \quad . \quad . \quad . \quad (35)$$

$$\text{curl } \boldsymbol{H}(\gamma) - j\omega(\gamma)\epsilon_d \boldsymbol{E}(\gamma) = 0 \quad . \quad . \quad . \quad (36)$$

When γ is imaginary $\omega(\gamma)$ is real and lies within a pass-band. In this case, taking the complex conjugate of eqns. (35) and (36) we obtain immediately

$$\text{curl } \boldsymbol{E}^*(\gamma) + j\omega(\gamma)\mu_d[-\boldsymbol{H}^*(\gamma)] = 0 \quad . \quad . \quad (37)$$

$$\text{curl } [-\boldsymbol{H}^*(\gamma)] - j\omega(\gamma)\epsilon_d \boldsymbol{E}^*(\gamma) = 0 \quad . \quad . \quad (38)$$

The complex conjugate field vectors are quasi-exponential functions with propagation constant γ^* which is equal to $-\gamma$ because γ is imaginary. Moreover $\boldsymbol{E}^*(\gamma)$ is tangential to the metal surfaces. Hence, if eqns. (37) and (38) are compared with eqns. (35) and (36), it is seen that, apart from a trivial constant factor, the complex vectors of the reversed wave with propagation coefficient $-\gamma$ are

$$\left.\begin{aligned} \boldsymbol{E}(-\gamma) &= \boldsymbol{E}^*(\gamma) \\ \boldsymbol{H}(-\gamma) &= -\boldsymbol{H}^*(\gamma) \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (39)$$

when γ is imaginary. Setting γ equal to Γ we obtain eqn. (13) in the text. It is obvious that this simple relation breaks down if γ is anything other than imaginary. However, two useful results can easily be obtained in which γ is not so restricted.

Let $\boldsymbol{E}(-\gamma)$ and $\boldsymbol{H}(-\gamma)$ be the complex vectors of the reversed wave with propagation coefficient $-\gamma$ and frequency $\omega(-\gamma)$. In the lossless dielectric these vectors satisfy Maxwell's equations

$$\text{curl } \boldsymbol{E}(-\gamma) + j\omega(-\gamma)\mu_d \boldsymbol{H}(-\gamma) = 0 \quad . \quad (40)$$

$$\text{curl } \boldsymbol{H}(-\gamma) - j\omega(-\gamma)\epsilon_d \boldsymbol{E}(-\gamma) = 0 \quad . \quad (41)$$

By taking the scalar product of eqns. (35), (36), (40) and (41) with $\pm \boldsymbol{H}(-\gamma)$, $-\boldsymbol{E}(-\gamma)$, $\boldsymbol{H}(\gamma)$ and $\mp \boldsymbol{E}(\gamma)$ respectively, adding and making use of the well-known formula for the divergence of a vector product, we can obtain

$$\begin{aligned} &\text{div } [\boldsymbol{E}(-\gamma) \times \boldsymbol{H}(\gamma) \pm \boldsymbol{E}(\gamma) \times \boldsymbol{H}(-\gamma)] \\ &\quad = -j[\omega(\gamma) \pm \omega(-\gamma)][\epsilon_d \boldsymbol{E}(-\gamma) \cdot \boldsymbol{E}(\gamma) \pm \mu_d \boldsymbol{H}(-\gamma) \cdot \boldsymbol{H}(\gamma)] \end{aligned}$$

$$. \quad . \quad . \quad . \quad (42)$$

in which either the upper or the lower signs can be taken together. This equation can be integrated over the region V of the dielectric bounded by the planes $z = 0$ and $z = D$ and the metal surfaces S. The volume integral of the divergence can be converted into a surface integral over the boundary which vanishes because of the perfectly conducting and quasi-exponential boundary conditions. Thus it is found that

$$[\omega(\gamma) \pm \omega(-\gamma)]\int_V [\epsilon_d \boldsymbol{E}(-\gamma).\boldsymbol{E}(\gamma) \pm \mu_d \boldsymbol{H}(-\gamma).\boldsymbol{H}(\gamma)]dV = 0 \quad . \quad . \quad . \quad . \quad (43)$$

If the lower signs are considered, it is found that

$$\omega(-\gamma) = \omega(\gamma) \quad . \quad . \quad . \quad . \quad . \quad (44)$$

i.e. the frequencies of the direct and reversed waves are the same. This follows because the integral certainly does not vanish. Indeed, when γ is imaginary it can be seen from eqn. (39) that it is just four times the mean energy stored in V. If the upper signs are considered and eqn. (44) is used, it is found immediately that

$$\int_V \epsilon_d \boldsymbol{E}(-\gamma).\boldsymbol{E}(\gamma)dV = -\int_V \mu_d \boldsymbol{H}(-\gamma).\boldsymbol{H}(\gamma)dV \quad . \quad (45)$$

This is a generalization of the well-known equality between the mean stored electric and magnetic energies to which it reduces when γ is imaginary, so that eqn. (39) is valid. Eqn. (20) is obtained by setting γ equal to Γ.

(8.2) Degenerate Modes

In this Appendix we shall derive the determinantal equation for the complex Q-factors of the various modes in the lossy guide which derive from a set of degenerate modes in the lossless guide.

Let $\boldsymbol{E}_l(\gamma_l)$ and $\boldsymbol{H}_l(\gamma_l)$ be the complex vectors of a mode in the lossy guide with propagation coefficient γ_l and frequency $\omega_l(\gamma_l)$. Suppose that it derives from a set of R degenerate modes in the lossless guide. Let $\boldsymbol{E}_r(-\gamma_l)$ and $\boldsymbol{H}_r(-\gamma_l)$, where r takes integral values from 1 to R inclusive, be the complex vectors of these modes, all with propagation coefficient $-\gamma_l$ and frequency $\omega(\gamma_l)$. Proceeding as in the text, R equations like eqn. (11) are obtained:

$$\frac{\omega_l(\gamma_l)}{2Q_c}\int_V [\epsilon_d \boldsymbol{E}_r(-\gamma_l) \cdot \boldsymbol{E}_l(\gamma_l) - \mu_d \boldsymbol{H}_r(-\gamma_l) \cdot \boldsymbol{H}_l(\gamma_l)]dV \quad (46)$$

$$= -\int_S \boldsymbol{E}_l(\gamma_l) \times \boldsymbol{H}_r(-\gamma_l) \cdot \boldsymbol{n}dS + \int_V \sigma_d \boldsymbol{E}_r(-\gamma_l) \cdot \boldsymbol{E}_l(\gamma_l)dV, \quad r = 1 \ldots R,$$

for the complex Q-factor Q_c of the mode considered. Here, as in eqn. (10),

$$\frac{\omega_l(\gamma_l)}{2Q_c} = j[\omega(\gamma_l) - \omega_l(\gamma_l)] \quad . \quad . \quad . \quad (47)$$

Suppose now that ω_l is given and that the losses are small. Then, with negligible error, γ_l in eqn. (46) can be replaced by the propagation coefficient Γ in the lossless guide at the given frequency ω_l. The tangential component of $\boldsymbol{E}_l(\Gamma)$ in the surface integral can then be approximated by the well-known formula

$$\boldsymbol{E}_{lt}(\Gamma) = \boldsymbol{H}_l(\Gamma) \times \boldsymbol{n}(1 + j)/\sigma_m\delta \quad . \quad . \quad . \quad (48)$$

Finally, the mode in the lossy guide is approximately equal to some linear combination of the degenerate modes in the lossless guide; i.e.

$$\left.\begin{aligned} \boldsymbol{E}_l(\Gamma) &\simeq \sum_{s=1}^{R} a_s \boldsymbol{E}_s(\Gamma) \\ \boldsymbol{H}_l(\Gamma) &\simeq \sum_{s=1}^{R} a_s \boldsymbol{H}_s(\Gamma) \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (49)$$

where the coefficients a_s remain to be determined. Thus from eqn. (46) we obtain

$$\sum_{r=1}^{R} D_{rs}a_s = 0, \quad r = 1 \ldots R \quad . \quad . \quad . \quad (50)$$

where

$$D_{rs} = \frac{\omega_l}{2Q_c}\int_V [\epsilon_d \boldsymbol{E}_r(-\Gamma).\boldsymbol{E}_s(\Gamma) - \mu_d \boldsymbol{H}_r(-\Gamma).\boldsymbol{H}_s(\Gamma)]dV$$

$$+ (1 + j)\int_S (\sigma_m\delta)^{-1}\boldsymbol{H}_r(-\Gamma).\boldsymbol{H}_s(\Gamma)dS - \int_V \sigma_d \boldsymbol{E}_r(-\Gamma).\boldsymbol{E}_s(\Gamma)dV \quad . \quad . \quad . \quad . \quad (51)$$

The set of homogeneous linear equations (50) for the coefficients a_s has a non-trivial solution only if the determinant of its matrix is zero, i.e.

$$|D_{rs}| = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (52)$$

The various integrals in eqn. (51) can be calculated if the fields in the lossless guide are known. Hence this determinantal equation can be solved to give the values of Q_c appropriate to the various modes of the lossy guide which derive from the degenerate modes of the lossless guide. The corresponding values of the coefficients a_s can be obtained subsequently from eqn. (50). So far as the propagation coefficients are concerned, however, the theory from this point onwards is identical to that in the text. Papadopoulos has given a more extensive treatment of degenerate modes in the case of uniform guides.[4]

621.371 : 621.396.674.1

Paper No. 2009 R
May 1956

AN INVESTIGATION OF THE PROPERTIES OF RADIAL CYLINDRICAL SURFACE WAVES LAUNCHED OVER FLAT REACTIVE SURFACES

By W. M. G. FERNANDO, B.Sc.(Eng.), Associate Member, and Professor H. E. M. BARLOW, Ph.D., B.Sc.(Eng.), Member.

(*The paper was first received 29th October, and in revised form 29th November*, 1955.)

SUMMARY

The problem of launching and supporting the radial form of surface wave over a flat surface is examined both theoretically and experimentally. The arrangement discussed consists of a vertical dipole aerial erected at different heights above the centre of a large horizontal metal plate inductively loaded at the surface, either by coating it with a thin layer of dielectric or by corrugating it with a series of concentric ridges. The earlier theoretical work of Cullen on the launching of a plane surface wave has been extended to the present case, and the results obtained have been substantially confirmed by the experimental observations recorded. In particular, the practicability of securing launching efficiencies approaching 80% for a particular height of aerial above a given surface has been fully demonstrated.

LIST OF PRINCIPAL SYMBOLS

r, ϕ and y = Cylindrical co-ordinates.
ρ, ϕ and θ = Spherical polar co-ordinates.
E_r and E_y = Electric-field components in the directions r and y.
H_ϕ = Magnetic field in the direction ϕ.
μ_0 and ϵ_0 = Permeability and permittivity of free space.
γ = Propagation coefficient of free space.
λ_0 = Free-space wavelength.
β = Phase-change coefficient.
$Z_s = R_s + jX_s$ = Surface impedance.
$H_n^{(1)}$ = Hankel function of the first kind and order n.
$H_n^{(2)}$ = Hankel function of the second kind and order n.
ω = Angular frequency.
Δ = Skin depth in metal boundary.
σ = Conductivity.
a = Thickness of dielectric coating or depth of corrugation.
D = Pitch of corrugation.
d = Width of slot forming corrugation.
h = Height of aerial above surface.
P_R = Power in radiation field.
P_S = Power in surface-wave field.
η = Launching efficiency of surface wave.
Π = Hertz vector.
F_r = Fresnel reflection coefficient.
θ = Angle of incidence of wave.

(1) INTRODUCTION

The form of wave which is now recognized as a 'surface wave' is characterized by its ability to propagate along an interface between two different media without radiation so that the only flow of energy normal to the interface is that required to supply the losses in the media concerned.[1] To comply precisely with these conditions the interface, in the direction of propagation of the wave, must be straight, but in the transverse plane it can take a variety of shapes. Thus we have the Zenneck wave[2] associated with a flat surface, which has been known for many years, although, until recently, only as a solution of Maxwell's equations. Then there is the more familiar axial cylindrical surface wave supported by a single-wire transmission line, which has been the subject of much recent work.[3,4,5,6]

The present investigation is concerned with the radial form of the Zenneck wave launched by a vertical line source, in air, at a frequency of about 10 000 Mc/s, over a flat reactive surface (see Fig. 1). The object was to examine the behaviour of this wave both in regard to field distribution and launching efficiency, when using a vertical dipole at different heights above the supporting surface, consisting of a large flat sheet of aluminium either dielectric coated or corrugated by a series of ridges concentric with the dipole. In accordance with the essential character of a surface wave the field must have an evanescent distribution over the wavefront, and in the particular case considered here it would be expected to suffer an exponential decay with height above the surface [see Fig. 1(*c*)]. In the radial direction along the surface a field variation according to a Hankel function is predicted. A theoretical study of the problem of launching a plane surface wave from a horizontal slot radiator over a flat surface of this kind was made by Cullen,[7] who concluded that when the slot was at a particular height above the surface a launching efficiency of the order of 80% was possible. He showed that, in these circumstances at large distances from the source, there should be a very considerable measure of cancellation of the radiation field near the surface by interference between the direct and reflected waves, thus ensuring that most of the power went into the surface wave itself. In the present work Cullen's analysis has been extended to cover the case dealt with experimentally of a vertical dipole radiator of finite dimensions instead of a narrow slot and used to propagate radially over a flat surface. The measurements show good agreement with the theoretical predictions both in regard to the field distributions and to the launching efficiency for different heights of the dipole radiator.

Since this investigation was completed, two papers have been published dealing with closely allied work. Rich[8] has made similar measurements on a plane surface wave, but as is to be expected in this case, boundary disturbances arose and may account for the somewhat lower launching efficiencies he obtained. Brick[9] has examined the problem in a rather different way, but his observations, so far as they can be compared, are in general agreement.

(2) THE SURFACE IMPEDANCE AND ITS EFFECT ON THE DISTRIBUTION OF THE SURFACE-WAVE FIELD

(2.1) Field equations

The field pattern of the surface wave with which this work is

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. Fernando was formerly in the Electrical Engineering Department, University College, London, and is now at the University of Ceylon.
Dr. Barlow is Professor of Electrical Engineering at University College, London.

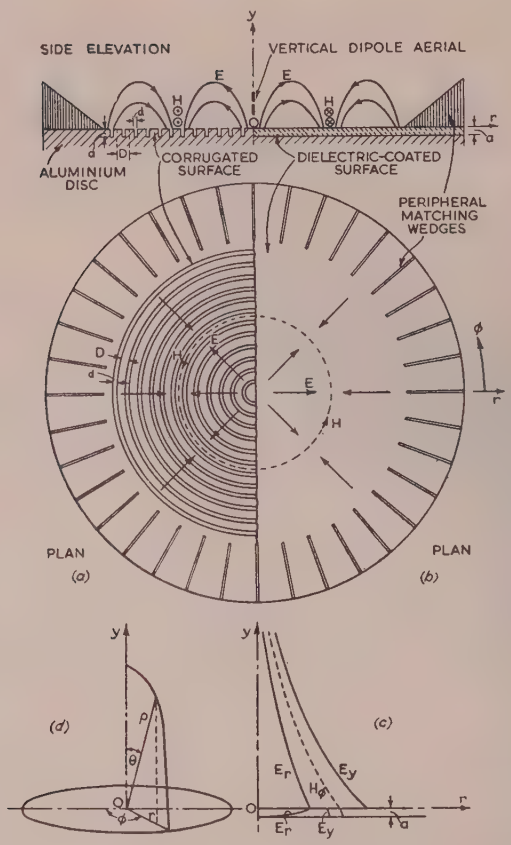


Fig. 1.—Loaded aluminium disc supporting a radial cylindrical surface wave.

(*a*) Corrugated surface.
(*b*) Dielectric-coated surface.
(*c*) Distribution of field components.
(*d*) Cylindrical and polar co-ordinates.

concerned is shown in Fig. 1, and in the air outside the surface ($y \geqslant 0$) its field components are given by

$$\left.\begin{aligned} H_\phi &= -A\varepsilon^{-\gamma_1 y}\mathrm{H}_1^{(2)}(j\beta_1 r) \\ E_r &= Z_s A\varepsilon^{-\gamma_1 y}\mathrm{H}_1^{(2)}(j\beta_1 r) \\ E_y &= \frac{\beta_1}{\gamma}\sqrt{\left(\frac{\mu_0}{\kappa_0}\right)}A\varepsilon^{-\gamma_1 y}\mathrm{H}_0^{(2)}(j\beta_1 r) \end{aligned}\right\} \quad . \quad . \quad . \quad (1)$$

where the factor $\varepsilon^{j\omega t}$ has been omitted for simplicity. $\gamma = \omega\sqrt{(\mu_0\epsilon_0)} = 2\pi/\lambda_0$.

$$\beta_1^2 - \gamma^2 + \gamma_1^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

and

$$Z_s = -\frac{E_r}{H_\phi} = \frac{j\gamma_1}{\gamma}\sqrt{\frac{\mu_0}{\epsilon_0}} = j\frac{\gamma_1}{\omega\epsilon_0} \quad . \quad . \quad . \quad . \quad (3)$$

Z_s, although independent of y, is actually defined for $y = 0$ as the surface impedance. Comparing eqn. (1) with the corresponding expressions for the field components of the plane Zenneck wave,[1,7] it will be seen that the rate of decay of the field with y, i.e. γ_1, is the same in the two cases, but along the surface the cylindrical form of wave falls off more rapidly owing to the expanding wavefront, and at large distances it varies as $\frac{1}{\sqrt{r}}\varepsilon^{-j\beta_1 r}$.

(2.2) Surface Impedance

In general, Z_s as defined by eqn. (3) for $y = 0$ is complex, so that

$$Z_s = R_s + jX_s \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

For homogeneous media of finite conductivity, R_s and X_s are physically inseparable since the existence of R_s necessarily implies a corresponding X_s quantity, arising from the penetration of the field. For a good conductor R_s and X_s are approximately equal, with R_s slightly the larger. When dealing with loss-free media it is possible to have X_s without R_s. Thus the effect of a thin dielectric coating on a smooth metal conductor is to enhance its surface reactance without appreciably affecting the surface resistance, and the reactance arising from the dielectric coating can be regarded as being directly added to the surface reactance of the bare metal.

For bare metal, we have

$$Z_s = \sqrt{\frac{j\omega\mu}{\sigma + j\omega\epsilon}} \simeq \sqrt{\frac{j\omega\mu}{\sigma}} = (1+j)\sqrt{\frac{\omega\mu}{2\sigma}} = R_s + jX_s \quad . \quad (5)$$

so that

$$R_s = X_s = \frac{\omega\mu\Delta}{2}$$

where the skin depth is

$$\Delta = \sqrt{\frac{2}{\omega\mu\sigma}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

From eqns. (3) and (4)

$$R_s + jX_s = j\frac{\gamma_1}{\omega\epsilon_0}$$

giving

$$\gamma_1 = \omega\epsilon_0(X_s - jR_s) \quad . \quad . \quad . \quad . \quad . \quad (7)$$

and at a frequency of 9 380 Mc/s, we find for a bare aluminium surface

$$\Delta = 0{\cdot}84 \times 10^{-6} \text{ metre}$$

$$X_s = R_s = 0{\cdot}03 \text{ ohm}$$

and

$$\gamma_1 = 0{\cdot}0168 - j0{\cdot}0168$$

The real part of γ_1, namely $\omega\epsilon_0 X_s$, determines the rate of decay of the field with distance from the surface, and the imaginary part, $\omega\epsilon_0 R_s$, is the associated phase coefficient. When $\omega\epsilon_0 R_s$ is large, the equi-phase surfaces, representing the wavefront, become appreciably inclined to the vertical, giving a corresponding component of power directed towards the surface to supply the losses in it.

For a flat surface to be capable of supporting a surface wave the decay coefficient $\omega\epsilon_0 X_s$ or the reactance X_s must be a finite positive quantity, and the larger its value the more closely does the field become concentrated in the immediate vicinity of the guiding surface. Whilst a smooth flat metal surface might be regarded as complying with the essential requirements, the reactance X_s and the corresponding decay coefficient are too small to make any surface wave in those circumstances identifiable without great difficulty. In the present work the reactance of the surface was therefore deliberately increased, either by coating it with a thin layer of dielectric or by corrugating it. With the use of a dielectric [see Fig. 1(*b*)] we get a stratified system. In the metal and in the air, the field decays exponentially with distance from the solid dielectric boundary, while within the solid dielectric itself we have a standing wave. It has been shown[1] that the impedance 'looking into' the surface of the solid dielectric is given by

$$Z_s = R_s + jX_s = \omega\mu_0\frac{\Delta}{2} + j\omega\mu_0\left[\left(\frac{\epsilon_r - 1}{\epsilon_r}\right)a + \frac{\Delta}{2}\right] \quad . \quad (8)$$

Since Δ is a relatively small quantity, the surface impedance in these circumstances becomes practically wholly reactive and of a value

$$Z_s \simeq jX_s = j\omega\mu_0 a\left(\frac{\epsilon_r - 1}{\epsilon_r}\right) \quad . \quad . \quad . \quad . \quad (9)$$

giving

$$\gamma_1 = -j\gamma\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}Z_s = \left(\frac{\epsilon_r - 1}{\epsilon_r}\right)\gamma^2 a \quad . \quad . \quad (10)$$

For a layer of Distrene ($\epsilon_r = 2{\cdot}56$) about $\frac{1}{16}$ in thick ($a = 0{\cdot}157$ cm) at a frequency of 9 380 Mc/s, we find

$$X_s = 70{\cdot}6 \text{ ohms}$$

$$R_s = 0{\cdot}03 \text{ ohm}$$

and

$$\gamma_1 = 0{\cdot}368 \text{ per centimetre}$$

A corrugated metal surface[5] [Fig. 1(*a*)] has a surface reactance X_s, given to a first approximation by

$$X_s = \frac{d}{D}\sqrt{\left(\frac{\mu_0}{\epsilon_0}\right)} \tan \gamma a \quad . \quad . \quad . \quad . \quad (11)$$

provided that there are at least three corrugations per wavelength along the surface. For slots milled out of the surface the resistive component of the impedance is relatively small and can be neglected.

Thus in this case,

$$\gamma_1 = -j\gamma\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}Z_s = \frac{d}{D}\gamma \tan \gamma a \quad . \quad . \quad . \quad (12)$$

For the frequency of 9 380 Mc/s, a slot depth $a = 0{\cdot}216$ cm, a slot width $d = 0{\cdot}305$ cm, and a pitch $D = 0{\cdot}75$ cm, we get $X_s = 69{\cdot}2$ ohms and $\gamma_1 = 0{\cdot}360$ per centimetre.

Although R_s for the corrugated surface is approximately double the value for the corresponding smooth surface, it is still negligible compared with X_s.

(3) EXPERIMENTAL APPARATUS

A large flat aluminium disc, 5 ft 6 in in diameter and $\frac{3}{8}$ in thick, was mounted with its surface horizontal on a steel trestle [Figs. 1(*a*) and 1(*b*)]. The surface was electrically loaded to enhance its reactance as described later. A glass rod projected through a hole drilled in the centre of the disc, and this rod, in supporting immediately above it the vertical dipole used for launching the wave, facilitated height adjustments of the aerial [see Fig. 4(*c*)]. In order to measure the radial distribution of the field at the surface, a slot $1\frac{1}{2}$ mm wide was cut in three sections, each of 24 cm length, along a radius in the aluminium disc so as to provide for a movable probe connected to a crystal detector. Arrangements were made to adjust the penetration of the probe and to ensure that it remained constant for a particular setting when the device was moved along the slot by its rack-and-pinion control (Fig. 2). To prevent a standing wave arising from reflections at the edge of the disc, a matched peripheral termination consisting of 72 triangular wedges of cardboard painted with colloidal graphite were used at 5° intervals around the circumference [Figs. 1(*a*) and 1(*b*)]. For measurement of the field at various points above the surface, a half-wave slot formed in the end of a short length of rectangular waveguide, tapered to reduce disturbance of the field, was used as a probe and associated with a crystal detector, as shown in Fig. 3.

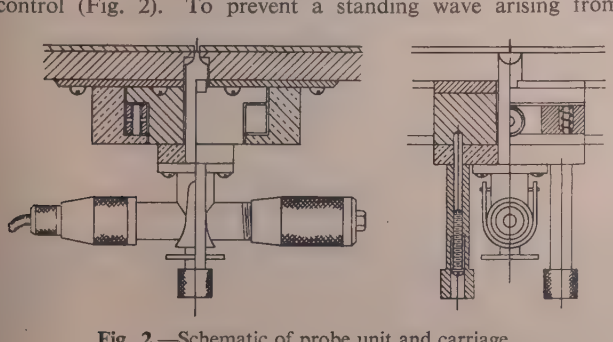


Fig. 2.—Schematic of probe unit and carriage.

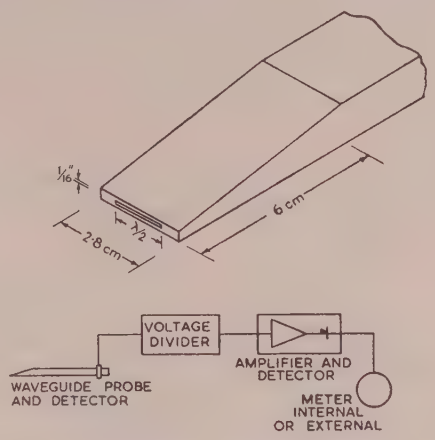


Fig. 3.—Waveguide probe and associated measuring equipment.
(*a*) Schematic of probe unit and carriage.
(*b*) Detector and measuring equipment.

This unit was mounted on a structure which rested on the aluminium disc and consisted of cardboard wedges painted with colloidal graphite similar to those of the peripheral matching system. Only the tip of the waveguide probe was allowed to project beyond the wedges into the field, so that most of the power incident on the probe extraneously was absorbed instead of being reflected by its various metal parts. For measurements in a vertical plane the probe unit was suspended by a threaded Distrene rod attached to the cardboard wedge structure and arranged to act as a traverse for small movements normal to the surface. A similar arrangement was used for observing the polar distribution of the field, except that the movement of the probe unit was in this case an angular one at a constant radius in a vertical plane. When wide differences of field strength were experienced over the range to be surveyed, it was found convenient to apply square-wave modulation at 3 kc/s to the klystron oscillator exciting the dipole aerial, and to employ an amplifier with a voltage divider between this and the crystal detector [Fig. 3(*b*)].

The loading of the aluminium disc to increase its surface reactance was carried out in the first place by covering it completely with a sheet of Distrene in ten sections $\frac{1}{16}$ in thick throughout (see Fig. 1(*b*)]. The Distrene was laid directly on the metal with adhesive tape holding together the different sections, whilst a radial slot was cut in it to correspond to the slot in the metal. The small air-gaps between the dielectric and the metal were not sufficient to cause any significant disturbance The measurements with the Distrene coating having been completed, it was removed and a series of 100 concentric corrugations of rectangular section were cut in the aluminium [see Fig. 1(*a*)]. The corrugations were 0·216 cm deep, 0·305 cm wide and had a pitch of 0·75 cm. The innermost groove had a radius of 4 cm, and to maintain the loading right up to the dipole aerial at the centre, a small disc of Distrene $\frac{1}{16}$ in thick and of 3 cm radius was placed on the surface around the dipole. The outermost corrugations were successively reduced in depth to zero at the edge

of the disc so as to avoid any serious electrical discontinuity. The approximate calculations in Section 2.2 for these loading arrangements indicate that they each give a surface reactance of about 70 ohms, and this was found to be suitable for the measurements planned. Some preliminary work with lighter loading showed that difficulties then arose owing to the spread of the field above the surface.

In making the choice of a vertical dipole aerial to launch the wave, consideration was given to the method of feeding the dipole and to alternative forms of radiator. A horizontal slot cut in the cylindrical surface of an E_{01} circular-section waveguide projecting vertically through a hole in the centre of the aluminium disc [Fig. 4(*a*)] would probably have been satisfactory, but it

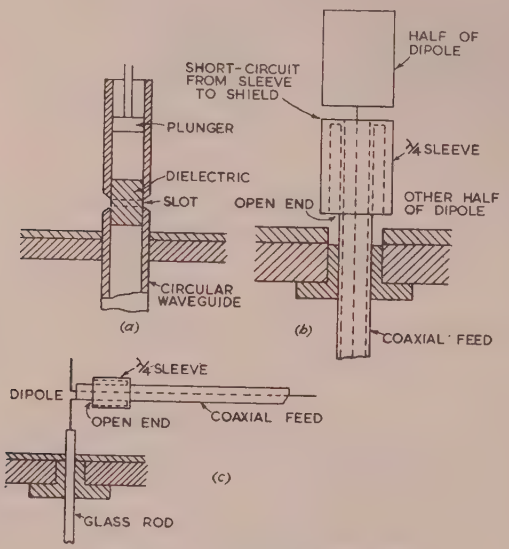


Fig. 4.—Launching arrangements.

was not tried because dipole radiators [Figs. 4(*b*) and 4(*c*)] had already been set up. In some experiments a vertical dipole, as nearly as possible balanced, was fed by a coaxial mineral-insulated cable from immediately beneath the disc [Fig. 4(*b*)]. This had the advantage of circular symmetry in a horizontal plane, but owing to some unavoidable electrical asymmetry of the dipole, it was found that stray currents flowed on the outside of the coaxial-cable feed giving a seriously distorted free-space polar diagram of radiation. To avoid this difficulty it was found preferable to feed the dipole by a coaxial cable above the surface and parallel with it, as shown in Fig. 4(*c*). A quarter-wave sleeve soldered on the outside of the feed cable provided a choke for any stray currents, and the radiation remote from the feed cable was then found to conform to the ideal pattern. The lack of circular symmetry was not of importance in these measurements.

(4) THEORETICAL CALCULATION FOR LAUNCHING EFFICIENCY OF THE SURFACE WAVE

In general, the total power radiated by an aerial located in air at a given height h above a horizontal flat surface can be regarded as being made up of two parts, P_R and P_S, associated respectively with the radiation and the surface-wave fields.

We can therefore define the launching efficiency of the surface wave as

$$\eta = \frac{P_S}{P_R + P_S} \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

To determine P_R and P_S we first obtain the magnetic com-

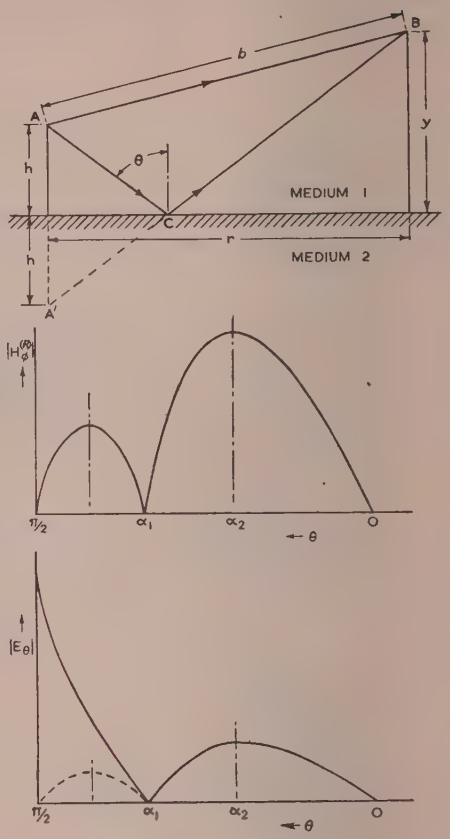


Fig. 5.—Components of field over the surface.

ponents of the radiation and surface-wave fields arising from a point source and assuming lossless media [see Fig. 5(*a*)]. Following the method of Sommerfeld for propagation over a flat earth we find that for radiation arriving by the *direct path* at a point at a height y above the surface, at a horizontal distance r and a radial distance b from the aerial, the Hertz vector is

$$\Pi = \frac{p}{4\pi\epsilon}\frac{\varepsilon^{-j\gamma b}}{b} = \frac{p}{4\pi\epsilon}\int_0^\infty \frac{J_0(\beta r)\varepsilon^{-y\sqrt{(\beta^2-\gamma^2)}}}{\sqrt{(\beta^2-\gamma^2)}}\beta d\beta \; . \quad (14)$$

where

$$p = \int_v P_0 dv \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

and P_0 is the moment per unit volume of the free-charge distribution at the elemental volume dv. The corresponding Hertz vector for the *reflected field* arriving by the indirect path at the same point (r, y) is given by

$$\Pi_{sec} = \frac{p}{4\pi\epsilon}\int_0^\infty \frac{F_r J_0(\beta r)\varepsilon^{-(h+y)\sqrt{(\beta^2-\gamma^2)}}}{\sqrt{(\beta^2-\gamma^2)}}\beta d\beta \; . \quad . \quad (16)$$

(4.1) Dielectric-coated Metal Surface

If we assume a smooth metal surface of infinite conductivity coated with a uniform thickness a of lossless dielectric having a relative permittivity ϵ_r, then for an angle of incidence θ measured from the vertical, the reflection coefficient of a plane or conical wave is found to be

$$F_r = \frac{j\gamma\cos\theta + \frac{\gamma}{\epsilon_r}\sqrt{(\epsilon_r - \sin^2\theta)}\tan\gamma a\sqrt{(\epsilon_r - \sin^2\theta)}}{j\gamma\cos\theta - \frac{\gamma}{\epsilon_r}\sqrt{(\epsilon_r - \sin^2\theta)}\tan\gamma a\sqrt{(\epsilon_r - \sin^2\theta)}} \quad . \quad (17)$$

or since

$$\beta = \gamma\sin\theta \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

we have

$$F_r = \frac{\epsilon_r\sqrt{(\beta^2 - \gamma^2)} + \sqrt{(\epsilon_r\gamma^2 - \beta^2)}\tan a\sqrt{(\epsilon_r\gamma^2 - \beta^2)}}{\epsilon_r\sqrt{(\beta^2 - \gamma^2)} - \sqrt{(\epsilon_r\gamma^2 - \beta^2)}\tan a\sqrt{(\epsilon_r\gamma^2 - \beta^2)}} \quad . \quad (19)$$

When a is small $\tan a\sqrt{(\epsilon_r\gamma^2 - \beta^2)} \simeq a\sqrt{(\epsilon_r\gamma^2 - \beta^2)}$ and this expression for F_r can be simplified.

(4.1.1) **Surface-Wave Field.**

To obtain the surface-wave field we evaluate the residue at the only pole which occurs when $\beta = \beta_1$ corresponding to an infinite value of the reflection coefficient F_r, which is the condition required for establishing a finite surface wave, and from eqn. (19)

$$\epsilon_r\sqrt{(\beta_1^2 - \gamma^2)} = \sqrt{(\epsilon_r\gamma^2 - \beta_1^2)}\tan a\sqrt{(\epsilon_r\gamma^2 - \beta_1^2)} \quad . \quad (20)$$

For $\beta \simeq \beta_1$, the denominator $D(\beta)$ of eqn. (19) can be expanded by Taylor's theorem in powers of $S = (\beta - \beta_1)$ to give

$$D(\beta) = D(\beta_1 + S) = D(\beta_1) + SD'(\beta_1) + \frac{S^2}{2!}D''(\beta_1) + \ldots$$

and since

$$D(\beta_1) = 0$$

we find near

$$\beta = \beta_1$$

that

$$D(\beta) = (\beta - \beta_1)D'(\beta_1) \quad . \quad . \quad . \quad . \quad (21)$$

Making use of the transformation

$$\int_0^\infty F(\beta)J_0(\beta r)\beta d\beta = \tfrac{1}{2}\int_{-\infty}^{+\infty} F(\beta)H_0^{(2)}(\beta r)\beta d\beta$$

where $F(\beta)$ is an even-valued function of β, we can rewrite eqn. (16) as follows:

$$\Pi_{sec} = \frac{p}{8\pi\epsilon}\int_{-\infty}^{+\infty} \frac{F_r\varepsilon^{-(h+y)\sqrt{(\beta^2-\gamma^2)}}H_0^{(2)}(\beta r)\beta d\beta}{\sqrt{(\beta^2 - \gamma^2)}} \quad . \quad (22)$$

$H_0^{(2)}(\beta r)$ being chosen to yield an outward travelling wave for real positive values[7] of β.

Using eqns. (19), (20) and (21) we obtain F_r near β_1 as

$$F_r = \frac{2\epsilon_r\sqrt{(\beta_1^2 - \gamma^2)}}{(\beta - \beta_1)D'(\beta_1)} \quad . \quad . \quad . \quad . \quad (23)$$

and

$$\Pi_{sec} = \frac{p\epsilon_r\beta_1}{4\pi\epsilon D'(\beta_1)}\varepsilon^{-(h+y)\sqrt{(\beta^2-\gamma^2)}}H_0^{(2)}(\beta_1 r)\int_c \frac{d\beta}{\beta - \beta_1} \quad . \quad (24)$$

where c is a small circle around β_1.

From Cauchy's theorem the contour integral is

$$\int_c \frac{d\beta}{\beta - \beta_1} = 2\pi j$$

and since the magnetic component of the surface-wave field $H_\phi^{(S)}$ is given by

$$H_\phi^{(S)} = \frac{\gamma^2}{j\omega\mu}\frac{\partial\Pi_{sec}}{\partial r} \quad . \quad . \quad . \quad . \quad . \quad (25)$$

we find, using the large argument approximation for the Hankel function, that

$$H_\phi^{(S)} = -p\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}\frac{\gamma\epsilon_r\beta_1^2}{\epsilon\sqrt{(2\pi\beta_1 r)}D'(\beta_1)}\varepsilon^{-j(\beta_1 r - 3\pi/4)}\varepsilon^{-\gamma_1(y+h)} \quad . \quad (26)$$

If instead of a point source the aerial consists of a vertical half-wave dipole with a current distributed sinusoidally and of value I at the centre, situated at height h above the surface, then eqn. (26) is modified as follows:

For an elemental dipole of length l

$$p = u\frac{j}{\omega}Il \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

where u is the unit vector in the direction of l. Thus integrating the contributions from elements of length δx of a dipole of finite dimensions $x = +\lambda/4$ to $-\lambda/4$ and having a current distribution $I(x) = I\cos\gamma x$ we get, after putting $I = 1$ for simplicity:

$$H_\phi^{(S)} = -\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}\frac{j2\gamma^2\epsilon_r}{\omega\epsilon\sqrt{(2\pi\beta_1 r)}D'(\beta_1)} \varepsilon^{-j(\beta_1 r - 3\pi/4)}\varepsilon^{-\gamma_1(y+h)}\cosh\left(\frac{\gamma_1\lambda}{4}\right) \quad . \quad (28)$$

(4.1.2) **Radiation Field.**

For the radiation field it is convenient to use spherical polar co-ordinates (ρ, θ, ϕ) instead of the cylindrical co-ordinates (r, y, ϕ), so that $\rho = \sqrt{(r^2 + y^2)}$ [see Fig. 1(*d*)]. Applying eqns. (14) and (16) for the direct and reflected fields, and remembering that for very large values of r and y we can legitimately use the stationary phase principle in respect of the slowly varying factor F_r, we find the *resultant* radiation field from a point source as

$$H_\phi^{(R)} = \frac{p\gamma^2}{4\pi\epsilon}\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}\frac{\varepsilon^{-j\gamma b}}{b}(\varepsilon^{j\gamma h\cos\theta} + F_r\varepsilon^{-j\gamma h\cos\theta})\sin\theta \quad . \quad (29)$$

Using eqn. (27) this expression is modified for a half-wave dipole with sinusoidal current distribution (putting $I = 1$) to

$$H_\phi^{(R)} = \frac{j2\gamma}{4\pi\omega\epsilon\sin\theta}\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}\frac{\varepsilon^{-j\gamma b}}{b} (\varepsilon^{j\gamma h\cos\theta} + F_r\varepsilon^{-j\gamma h\cos\theta})\cos\left(\frac{\pi\cos\theta}{2}\right) \quad . \quad (30)$$

(4.1.3) **Power in the Surface-Wave Field.**

Neglecting the power (about 6% of the surface-wave power) in the thin layer of dielectric over the metal surface, we can write for the surface-wave power

$$P_S = \int_0^\infty |H_\phi^{(S)}|^2\sqrt{\left(\frac{\mu_0}{\epsilon_0}\right)}\frac{\beta_1}{\gamma}2\pi r dy \quad . \quad . \quad . \quad (31)$$

Thus for a point-source radiator we get

$$P_S = \frac{\gamma}{2\gamma_1}\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}\left[\frac{p\epsilon_r\beta_1^2}{\epsilon D'(\beta_1)}\right]^2\varepsilon^{-2\gamma_1 h} \quad . \quad . \quad . \quad (32)$$

or, for a half-wave dipole with sinusoidal current distribution,

$$P_S = \frac{2\gamma^3\epsilon_r^2}{\gamma_1}\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}\left[\frac{\cosh\left(\frac{\gamma_1\lambda}{4}\right)}{\omega\epsilon D'(\beta_1)}\right]^2\varepsilon^{-2\gamma_1 h} \quad . \quad . \quad (33)$$

Values of P_S obtained from these expressions for different aerial heights are plotted in Fig. 6.

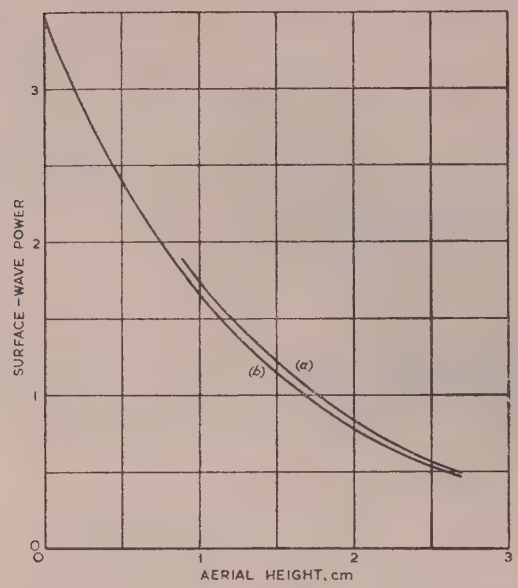


Fig. 6.—Variation of surface-wave power with aerial height for coated surface.

(*a*) Half-wave dipole. (*b*) Point source.

(4.1.4) **Power in the Radiation Field.**

Similarly the power in the radiation field is

$$P_R = \int_0^{\pi/2} |H_\phi^{(R)}|^2 \sqrt{\left(\frac{\mu_0}{\epsilon_0}\right)} 2\pi b^2 \sin\theta d\theta \quad . \quad . \quad . \quad (34)$$

and for a point-source radiator we therefore have

$$P_R = \frac{p^2\gamma^4}{8\pi\epsilon^2}\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)} \int_0^{\pi/2} \mathrm{F}^2(\theta) \sin^3\theta d\theta \quad . \quad . \quad . \quad (35)$$

where $$\mathrm{F}(\theta) = |\varepsilon^{j\gamma h\cos\theta} + F_r\varepsilon^{-j\gamma h\cos\theta}| \quad . \quad . \quad . \quad (36)$$

Using the value of F_r, as given by eqn. (19) when a is small, and putting

$$\Phi = \arctan\left[\frac{\cos\theta}{\gamma a\left(\frac{\epsilon_r - 1}{\epsilon_r} + \frac{\cos^2\theta}{\epsilon_r}\right)}\right] \quad . \quad . \quad (37)$$

we have a convenient form for evaluating $\mathrm{F}^2(\theta)$ as

$$\mathrm{F}^2(\theta) = 2 - 2\cos(2\Phi - 2\gamma h\cos\theta) \quad . \quad . \quad (38)$$

and we can then obtain P_R for a given aerial height h by numerical integration using Simpson's rule. For that purpose $\mathrm{F}^2(\theta)\sin^3\theta$ was calculated for $\theta = 0°$ to 90° at intervals of 5°, and the resulting curve of P_R is shown in Fig. 7. For a half-wave dipole aerial, putting $I = 1$, the corresponding expression for the radiated power is

$$P_R = \frac{\gamma^2}{2\pi\omega^2\epsilon^2}\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)} \int_0^{\pi/2} \frac{\mathrm{F}^2(\theta)}{\sin\theta} \cos^2\left(\frac{\pi\cos\theta}{2}\right) d\theta \quad . \quad (39)$$

Values obtained from this expression by numerical integration are also given in Fig. 7.

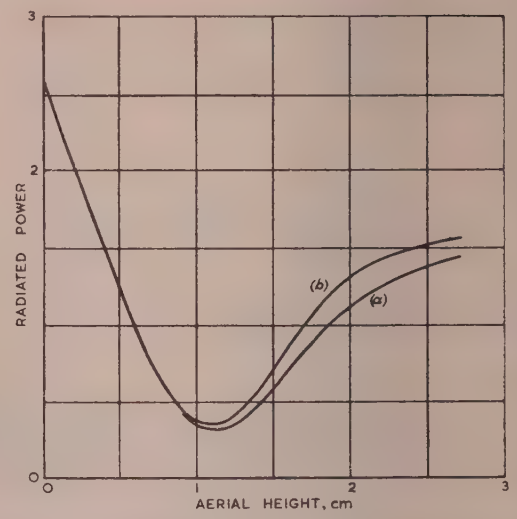


Fig. 7.—Variation of radiated power with aerial height for coated surface.

(*a*) Half-wave dipole. (*b*) Point source.

(4.1.5) **Launching Efficiency for the Surface Wave over a Dielectric-Coated Surface.**

Having evaluated P_R and P_S for various values of h, the launching efficiency η for the surface wave can be found immediately from eqn. (13). In Fig. 8 curves of η are given for both

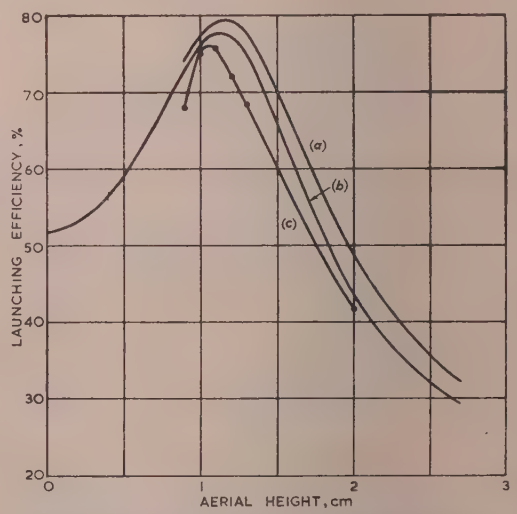


Fig. 8.—Launching efficiency for coated surface.

(*a*) Half-wave dipole. (*b*) Point source.
(*c*) Experimental results.

point-source and half-wave dipole aerials, showing that a maximum efficiency of over 80% is to be expected at an aerial height of about 1 cm.

(4.2) Corrugated Metal Surface

Expressions for the field excited over a corrugated metal surface, in which the losses are neglected, can be obtained by much the same procedure as for the dielectric-coated surface.

The reflection coefficient for the corrugated surface is given by[7]

$$F_r = \frac{1 - jX_s\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}\sec\theta}{1 + jX_s\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}\sec\theta} \quad . \quad . \quad . \quad . \quad (40)$$

and from eqn. (10) when $R_s = 0$

$$\gamma_1 = \gamma X_s\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}$$

so that, using eqn. (18), we find

$$F_r = \frac{\sqrt{(\beta^2 - \gamma^2)} + \gamma_1}{\sqrt{(\beta^2 - \gamma^2)} - \gamma_1} \quad . \quad . \quad . \quad . \quad (41)$$

(4.2.1) Surface-Wave Field.

From eqn. (41) for F_r the only pole occurs at $\beta = \beta_1$ where $\beta_1^2 - \gamma^2 = \gamma_1^2$.

Representing the denominator of eqn. (41) by $D(\beta)$ we get the value of F_r near $\beta = \beta_1$ as

$$F_r \doteq \frac{2\sqrt{(\beta_1^2 - \gamma^2)}}{(\beta - \beta_1)D'(\beta_1)}$$

and

$$\Pi_{sec} = \frac{jp\gamma_1}{2\epsilon}\varepsilon^{-\gamma_1(y+h)}H_0^{(2)}(\beta_1 r) \quad . \quad . \quad . \quad . \quad (42)$$

Thus for a point source

$$H_\phi^{(S)} = \frac{\gamma^2}{j\omega\mu}\frac{\partial\Pi_{sec}}{\partial r} = -p\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}\frac{\gamma\gamma_1}{\epsilon}\sqrt{\left(\frac{\beta_1}{2\pi r}\right)}\varepsilon^{-j(\beta_1 r - 3\pi/4)}\varepsilon^{-\gamma_1(y+h)} \quad . \quad . \quad . \quad . \quad (43)$$

after introducing the large-argument approximation for the Hankel function.

For a half-wave dipole with sinusoidal current distribution the corresponding expression (putting $I = 1$) is

$$H_\phi^{(S)} = -\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}\frac{j2\gamma^2\gamma_1}{\omega\epsilon\beta_1^2}\sqrt{\left(\frac{\beta_1}{2\pi r}\right)} \varepsilon^{-j(\beta_1 r - 3\pi/4)}\varepsilon^{-\gamma_1(y+h)}\cosh\left(\frac{\gamma_1\lambda}{4}\right) \quad . \quad (44)$$

(4.2.2) Radiation Field.

The procedure in determining $H_\phi^{(R)}$ is exactly the same as for the dielectric-coated surface (see Section 4.1.2) and we find the same expressions, namely eqn. (29), for a point source and eqn. (30) for a half-wave dipole aerial, but the reflection coefficient is now given by eqn. (41).

(4.2.3) Power in the Surface-Wave Field

Using eqn. (31), we get for a point source

$$P_S = \frac{\gamma\gamma_1}{2}\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}\left(\frac{p\beta_1}{\epsilon}\right)^2\varepsilon^{-2\gamma_1 h} \quad . \quad . \quad . \quad (45)$$

and for the half-wave dipole radiator,

$$P_S = 2\gamma^3\gamma_1\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}\left[\frac{\cosh\left(\frac{\gamma_1\lambda}{4}\right)}{\omega\epsilon\beta_1}\right]^2\varepsilon^{-2\gamma_1 h} \quad . \quad . \quad (46)$$

Calculated values for different aerial heights are shown in Fig. 9.

(4.2.4) Power in the Radiation Field.

Since the expressions for $H_\phi^{(R)}$ are the same as for the dielectric-coated surface, it follows that the P_R, when employing a corrugated metal surface, is given by eqn. (35) for a point-source radiator and by eqn. (39) for the half-wave dipole. In these expressions eqn. (36) gives $F(\theta)$ and eqn. (41) gives the reflection coefficient F_r.

Fig. 10 gives curves of P_R calculated from these expressions.

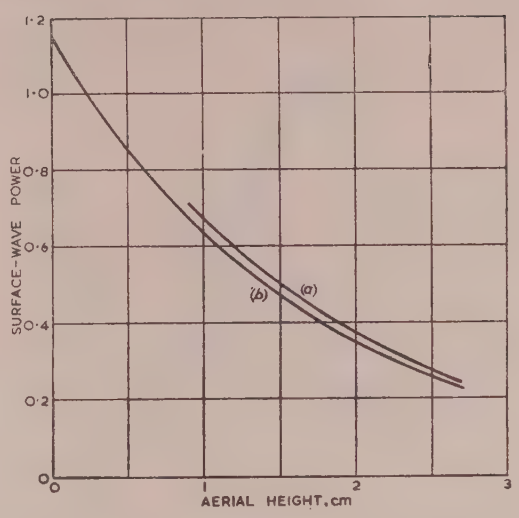


Fig. 9.—Variation of surface-wave power with aerial height for corrugated surface.

(*a*) Half-wave dipole. (*b*) Point source.

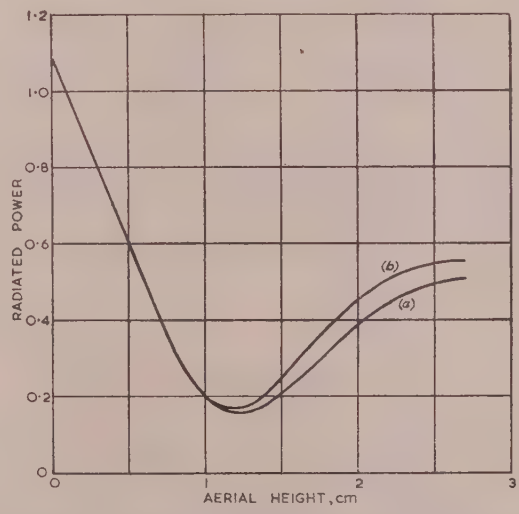


Fig. 10.—Variation of radiated power with aerial height for corrugated surface.

(*a*) Half-wave dipole. (*b*) Point source.

(4.2.5) Launching Efficiency for the Surface Wave Over a Corrugated Surface.

Using eqn. (13) and the appropriate values of P_R and P_S, we find the launching efficiency η for the surface wave, and the results for various aerial heights are shown in Fig. 11. For the corrugated surface the maximum launching efficiency appears to be rather lower than for the dielectric-coated surface, but a direct comparison between the two sets of curves is not possible since the values of γ_1 are not quite the same.

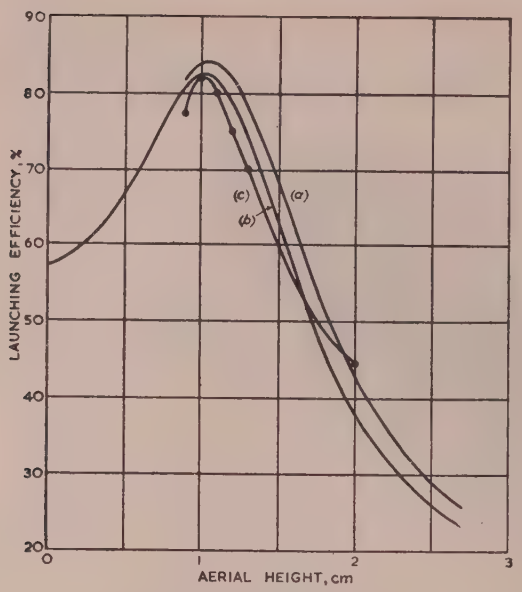


Fig. 11.—Launching efficiency for corrugated surface.
(*a*) Half-wave dipole. (*b*) Point source.
(*c*) Experimental results.

(5) EXPERIMENTAL OBSERVATIONS AND COMPARISON WITH THEORY

Before making measurements of the field distribution in conditions that would be expected to provide for a surface wave, some preliminary calibration experiments were carried out. In the first place the crystal detectors were tested using a waveguide test bench and found to obey accurately a square law over the range required. Secondly, the effectiveness of the matched termination at the periphery of the disc was checked by ensuring that there was no radial standing wave of significant amplitude.

Measurements were then made of the electric field strength at different distances from the dipole radiator and adjacent to the smooth flat surface of bare aluminium.

A logarithmic plot of this graph (Fig. 12) showed that the

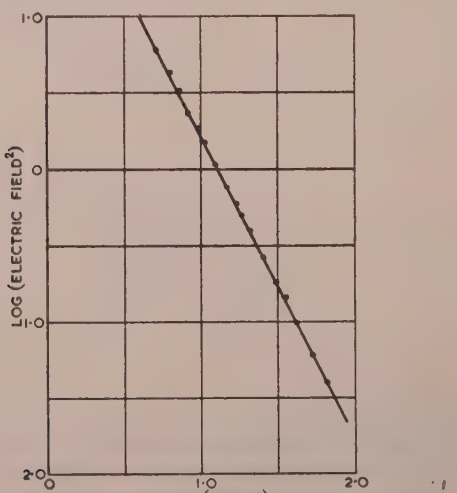


Fig. 12.—Logarithmic plot of radial field distribution over bare metal surface.

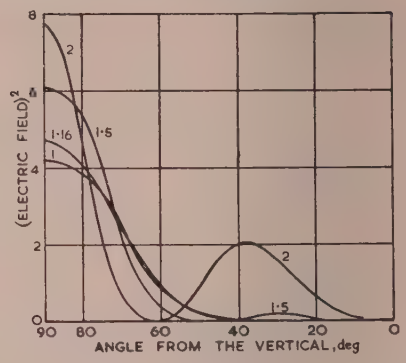


Fig. 13.—Radiation patterns of dipole used over bare metal surface.

field was practically inversely proportional to the radius—a result which was to be expected since no appreciable surface wave was likely to be excited in this case. The corresponding polar curve is shown in Fig. 13, and again this agrees with expectations assuming a pure radiation field.

(5.1) Dielectric-Coated Surface

(5.1.1) **Field Distribution.**

The next step was to repeat the measurements when the metal surface was coated with Distrene $\frac{1}{16}$ in thick, the field strength in these circumstances being shown in Fig. 14. The field

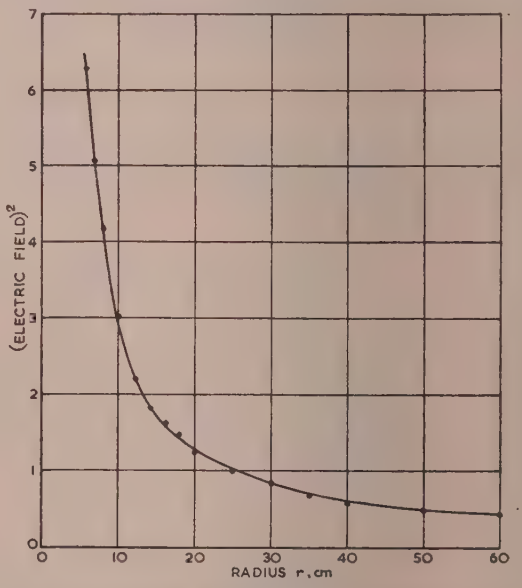


Fig. 14.—Radial field distribution over coated metal surface.

at ranges exceeding 10 cm is now found to be inversely proportional to the square root of the radius, and this conforms to the large-argument approximation for the Hankel function representing the surface wave. It therefore seems that in proximity to the highly reactive surface employed in these experiments the radiation field is small enough by comparison with the surface-wave field to be ignored. The field at a given horizontal radius and at different heights y up to 4 cm above the surface is recorded in Fig. 15, and by plotting the corresponding Naperian logarithm of the electric field, the decay coefficients in Table 1 were deduced.

In Section 2.2 the decay coefficient for this arrangement was

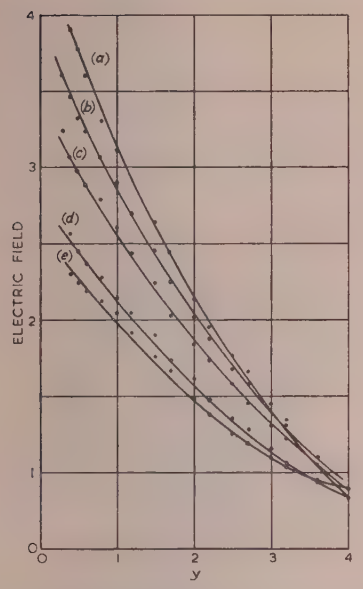


Fig. 15.—Vertical field distribution over coated surface for aerial height of 1·5 cm and values of r between 18 and 35 cm.
(*a*) $r = 18$ cm. (*b*) $r = 20$ cm. (*c*) $r = 24$ cm. (*d*) $r = 30{\cdot}6$ cm. (*e*) $r = 35$ cm.

Table 1

VERTICAL DECAY COEFFICIENT FOR THE SURFACE-WAVE FIELD

Aerial height	Radius	Decay coefficient
cm	cm	γ_1 per cm
1·5	18	0·374
1·5	20	0·359
1·5	24	0·346
1·5	30·5	0·326
1·5	35	0·296

calculated and found to be approximately 0·368 per centimetre—a value which agrees, within the limits of error, with the measurements. The assumption that the radiation field is negligible in these circumstances near the surface is justifiable, but a calculation for the total field, especially for larger values of y, is of interest. This can be done following Cullen,[7] who gives for the two-dimensional case

$$H_z = -2\gamma V\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}\left[\frac{\gamma_1}{\beta_1}\varepsilon^{-(\gamma_1 h+\gamma_1 y+j\beta_1 x)} + \frac{\varepsilon^{-j(\gamma x+\pi/4)}}{\sqrt{2\pi}}\left\{\frac{F_1(y_1h)}{(\gamma x)^{3/2}} - \frac{j3}{(\gamma x)^{5/2}}\left[\frac{F_1(y_1h)}{8}+F_2(y_1h)\right]+\cdots\right\}\right] \quad . \quad . \quad . \quad . \quad (47)$$

where V is the voltage across the launching aperture.

$$F_1(y_1h) = \frac{\gamma^2}{\gamma_1^2}(1-\gamma_1 y)(1-\gamma_1 h)$$

$$F_2(y_1h) = \frac{\gamma^4}{\gamma_1^4}\left[1-\gamma_1(y+h)+\frac{\gamma_1^2}{2}(y+h)^2 - \frac{\gamma_1^3}{6}(y+h)^3+\frac{\gamma_1^4}{6}yh(y^2+h^2)\right]$$

and provided that $x \gg (y+h)$ with $\gamma x \gg \left|\frac{\gamma}{\gamma_1}\right|^2$. If in eqn. (47) r is substituted for x we can apply the expression to calculate H_ϕ in the three-dimensional case. This has been done to evaluate $|H_\phi|$ as shown in Fig. 16, for aerial heights of 1 cm and 1·5 cm

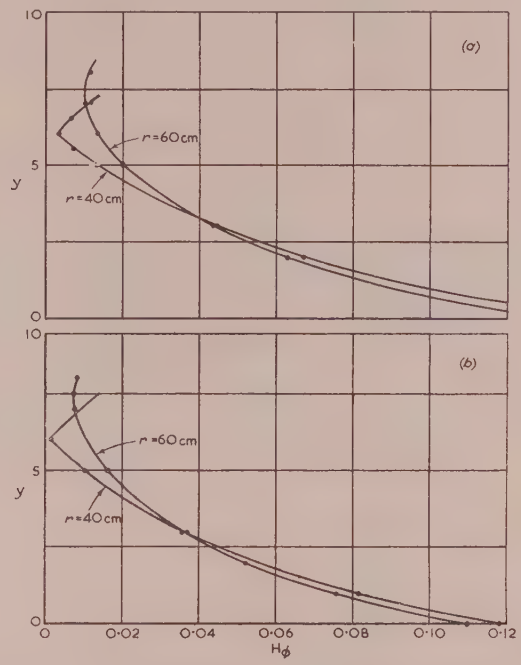


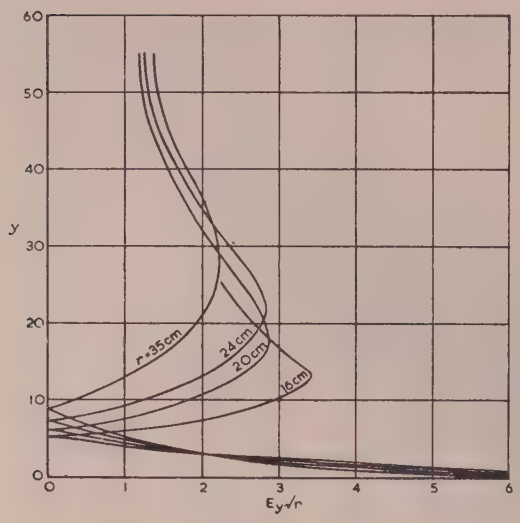
Fig. 16.—Calculated vertical field distribution including radiation component.
(*a*) For an aerial height of 1 cm with $r = 40$ cm and $r = 60$ cm.
(*b*) For an aerial height of 1·5 cm with $r = 40$ cm and $r = 60$ cm.

at horizontal distances from the aerial of 40 and 60 cm and for values of y up to 8 cm. Although some of the conditions are different, the general form of this curve is similar to the experimental one in Fig. 15. According to Cullen, it is to be expected that the radiation field vanishes at least to the order $1/(\gamma r)^{3/2}$ when $y = 1/\gamma_1$. This result is independent of θ when the angle of elevation $(\pi/2-\theta)$ is small, and arises from interference between the direct and reflected waves into which the radiation field can be resolved.

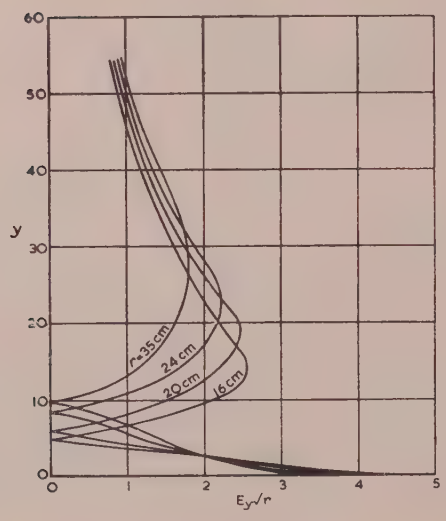
For the dielectric-loaded surface we have already found that $1/\gamma_1 = 2{\cdot}7$ cm, so that in the horizontal plane at $y = 2{\cdot}7$ cm the field should be entirely in the form of a surface wave. To examine this condition a series of measurements were made of the electric field E_y at different distances y from the surface, for horizontal radii r lying between 16 and 35 cm and two different aerial heights, namely 1 and 1·3 cm. The graphs [Figs. 17(*a*) and 17(*b*)], giving $E_y\sqrt{r}$ as a function of y, were plotted from these figures, and it will be observed that in one case the family of curves intersect approximately at a point $y = 3$ cm and in the other at $y = 2{\cdot}5$ cm. This signifies that, for the field concerned, $E_y\sqrt{r}$ becomes independent of r or E_y becomes proportional to $1/\sqrt{r}$ at the point of intersection, and that is characteristic of the pure surface wave. We conclude, therefore, that the theoretical prediction is, in fact, borne out by experiment to the degree of accuracy obtainable in these measurements.

(5.1.2) Power in Radiation and Surface-Wave Fields.

Eqn. (30) gives the magnetic component of the *radiation* field $H_\phi^{(R)}$ in the direction θ from the vertical, when excited by a half-



(*a*)



(*b*)

Fig. 17.—Plot of $E_y\sqrt{r}$ against y.
(*a*) For an aerial height of 1·3 cm and values of r between 16 and 35 cm.
(*b*) For an aerial height of 1 cm and values of r between 16 and 35 cm.

wave dipole. At a given frequency and at a constant radius b from the source we can write

$$|H_\phi^{(R)}| \propto \frac{F(\theta)\cos(\pi/2\cos\theta)}{\sin\theta} \quad . \quad . \quad . \quad . \quad (48)$$

where $F(\theta)$ is defined by eqn. (36).

In Fig. 18 curves calculated from eqn. (48) and giving $|H_\phi^{(R)}|$ as a function of θ for several different aerial heights h ranging from 0·9 to 2 cm are plotted. The corresponding *total* field distribution, which was conveniently measured in terms of $|E_\theta|$, is shown in Fig. 5(*c*); and comparing this experimentally derived curve for a particular value of h with the theoretical one redrawn in Fig. 5(*b*) representing the radiation field only, it was observed that, in both cases, the field fell to zero at $\theta = 0$ and again at $\theta = \alpha_1$, rising to a maximum at $\theta = \alpha_2$. It seemed justifiable to conclude, therefore, that the part of the curve representing *total* field between $\theta = 0$ and $\theta = \alpha_1$ in Fig. 5(*c*) does not include

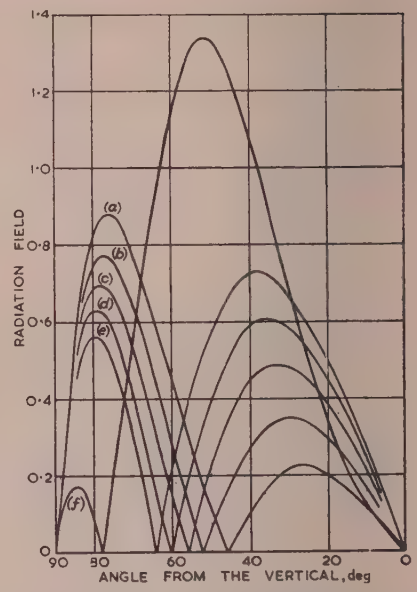


Fig. 18.—Calculated polar patterns of radiation for a half-wave dipole over a coated surface and the following aerial heights:
(*a*) 0·9 cm. (*d*) 1·2 cm.
(*b*) 1·0 cm. (*e*) 1·3 cm.
(*c*) 1·1 cm. (*f*) 2 cm.

any significant surface wave, but that the other part lying between $\theta = \alpha_1$ and $\theta = \pi/2$ represents a combination of both radiation and surface-wave fields. We require to find the radiation field by itself throughout the angle $\theta = 0$ to $\pi/2$, in order to deduce the power P_R in this field, and we therefore rely upon the theoretical calculation for the curve of radiation field between $\theta = \alpha_1$ and $\pi/2$, drawing this curve as shown dotted in Fig. 5(*c*) by fitting the rest of it to the corresponding curve in Fig. 5(*b*).

P_R is then obtained from

$$P_R = 2\pi b^2\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}\int_0^{\pi/2}|E_\theta^{(R)}|^2\sin\theta d\theta \quad . \quad . \quad . \quad (49)$$

by numerical integration using Simpson's rule.

To find the power P_S in the surface wave it is only necessary to know its field strength at one point, preferably very close to the supporting surface. Thus if the surface wave field at a point distant y_1 from the surface is $E^{(S)}(y_1)$, then for the field at any other point y we have

$$E^{(S)}(y) = E^{(S)}(y_1)\varepsilon^{\gamma_1(y_1-y)}$$

and hence

$$P_S = \int_0^\infty \frac{2\pi b\gamma}{\beta_1}|E^{(S)}(y)|^2\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}dy = \frac{\pi b\gamma}{\gamma_1\beta_1}\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}|E^{(S)}(y_1)|^2\varepsilon^{2\gamma_1 y_1}$$
$$. \quad . \quad . \quad . \quad (50)$$

(5.1.3) Launching Efficiency for Surface Wave.

Using values of P_R and P_S derived as explained in the preceding Section, the launching efficiency η was found from eqn. (13) and plotted for different aerial heights h as shown in Fig. 8.

(5.2) Corrugated Metal Surface

(5.2.1) Field Distribution.

The experimental procedure applied to the dielectric-coated surface was repeated almost exactly in making observations on the field above the corrugated supporting surface. In determin-

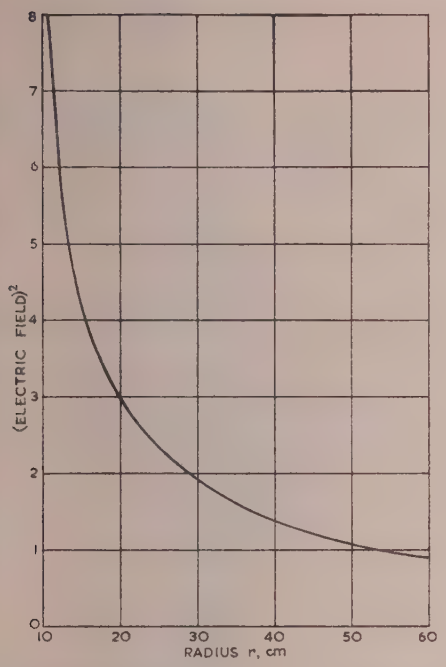


Fig. 19.—Radial field distribution over a corrugated surface for an aerial height of 1·3 cm.

ing the radial decay coefficient, care was taken to avoid field distortion in the neighbourhood of the probe of the detector unit by positioning it at the centre of a ridge in the corrugations while readings were being taken. Fig. 19 shows the measured values of the field at the surface for different distances r from the aerial. A logarithmic plot of this curve demonstrates that the field varies as $1/r^{0\cdot54}$ or roughly inversely as the square root of the radius, which is in accordance with a surface-wave distribution.

Fig. 20 gives the measured field distribution in a vertical plane

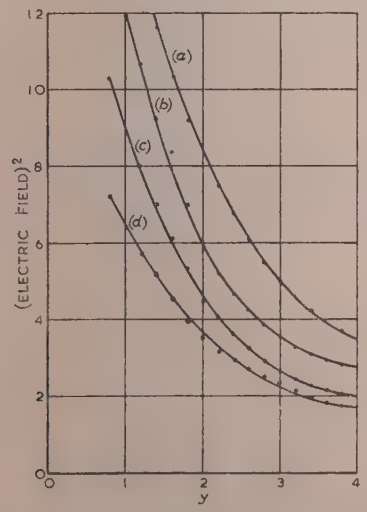


Fig. 20.—Vertical field distribution over a corrugated surface for an aerial height of 1·1 cm.
(*a*) $r = 25$ cm. (*c*) $r = 38$ cm.
(*b*) $r = 30$ cm. (*d*) $r = 45$ cm.

for different radii r, and a logarithmic plot of these curves yields the decay coefficients in Table 2.

Table 2

Aerial height	Radius	Decay coefficient
cm	cm	γ_1 per cm
1·1	25	0·267
1·1	30	0·310
1·1	38	0·310
1·1	45	0·292

The mean value of $\gamma_1 = 0\cdot3$ derived from these measurements was used in the calculations in preference to the theoretical figure of 0·36. A wavelength measurement along the surface confirmed the choice of $\gamma_1 = 0\cdot3$.

(5.2.2) **Power in Radiation and Surface-Wave Fields. Launching Efficiency.**

Using eqns. (48), (49) and (50) in exactly the same way as for the dielectric-coated surface, Fig. 21 was obtained, and experimental values for launching efficiency are shown in Fig. 11.

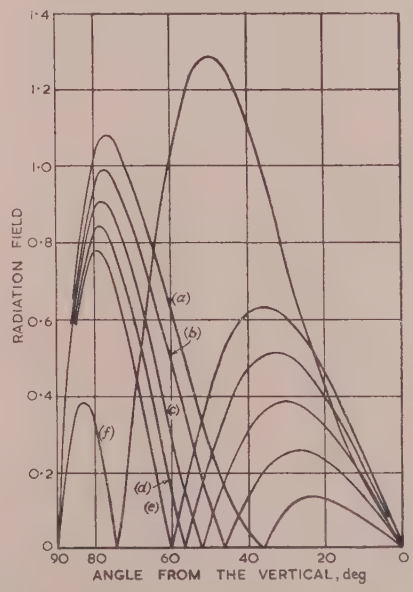


Fig. 21.—Calculated polar patterns of radiation for a half-wave dipole over a corrugated surface and the following aerial heights:
(*a*) 0·9 cm. (*b*) 1·0 cm. (*c*) 1·1 cm. (*d*) 1·2 cm. (*e*) 1·3 cm. (*f*) 2 cm.

(6) CONCLUSIONS

The experiments described demonstrate the practicability of launching and supporting a radial cylindrical surface wave over a highly reactive surface consisting of a metal plate either coated with a thin layer of dielectric or corrugated. The circular symmetry of this wave made possible a much more precise examination of its behaviour than could have been achieved in the case of the corresponding plane wave. The field distribution and launching efficiency calculated theoretically have been found to agree, within the limits of error, with the experimental observations. The high launching efficiency for a particular height of the aerial above the surface, as predicted by Cullen for a plane wave, has been substantially confirmed.

(7) ACKNOWLEDGMENTS

The work described was carried out as part of the programme of the Radio Research Board. The paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research. The authors wish to thank Prof. A. L. Cullen and Mr. H. G. Effemey for many helpful discussions in connection with this work and to acknowledge the valuable contribution of Mr. A. D. Woodward in the construction of the experimental apparatus. They are also grateful to the University of Ceylon and to the Department of Scientific and Industrial Research for financial support of the work.

(8) REFERENCES

(1) BARLOW, H. E. M., and CULLEN, A. L.: 'Surface Waves', *Proceedings I.E.E.*, Paper No. 1482 R, April, 1953 (**100,** Part III, p. 329).

(2) ZENNECK, J.: 'Uber die Fortpflanzung ebener elektromagnetischer Wellen langs einer ebener Leiterflache und ihre Beziehung zur drahtlosen Telegraphie', *Annalen der Physik*, 1907, **23,** p. 846.

(3) GOUBAU, G.: 'On the Excitation of Surface Waves', *Proceedings of the Institute of Radio Engineers*, 1952, **40,** p. 865.

(4) BARLOW, H. E. M., and KARBOWIAK, A. E.: 'An Investigation of the Characteristics of Cylindrical Surface Waves', *Proceedings I.E.E.*, Paper No. 1462 R, April, 1953 (**100,** Part III, p. 321).

(5) BARLOW, H. E. M., and KARBOWIAK, A. E.: 'An Experimental Investigation of the Properties of Corrugated Cylindrical Surface Waveguides', *ibid.*, Paper No. 1625 R, May, 1954 (**101,** Part III, p. 182).

(6) BARLOW, H. E. M., and KARBOWIAK, A. E.: 'An Experimental Investigation of Axial Cylindrical Surface Waves supported by Capacitive Surfaces', *ibid.*, Paper No. 1786 R, May, 1955 (**102** B, p. 313).

(7) CULLEN, A. L.: 'The Excitation of Plane Surface Waves', *ibid.*, Monograph No. 93 R, February, 1954 (**101,** Part IV, p. 225).

(8) RICH, G. J.: 'The Launching of a Plane Surface Wave', *ibid.*, Paper No. 1783 R, March, 1955 (**102** B, p. 237).

(9) BRICK, D. B.: 'The Excitation of Surface Waves by a Vertical Antenna', *Proceedings of the Institute of Radio Engineers*, 1955, **43,** p. 721.

(10) BROWN, J.: 'The Types of Wave which may exist near a Guiding Surface', *Proceedings I.E.E.*, Paper No. 1567 R, November, 1953 (**100,** Part III, p. 363).

621.372.2 : 621.317.6 : 621.317.729

Paper No. 1991 R
May 1956

AN INVESTIGATION INTO SOME FUNDAMENTAL PROPERTIES OF STRIP TRANSMISSION LINES WITH THE AID OF AN ELECTROLYTIC TANK

By J. M. C. DUKES, M.A., Associate Member.

(*The paper was first received 2nd August, and in revised form 12th November, 1955.*)

SUMMARY

The paper describes a method of deriving numerical values for quantities such as characteristic impedance, attenuation and wavelength with the aid of an electrolytic tank. Numerical results are given for two principal forms of strip transmission line for a range of dimensions for which rigorous solutions by direct analysis are not readily available. This range is nevertheless of considerable practical interest as it encompasses characteristic impedances between 20 and 150 ohms. The recent findings of Black and Higgins in regard to the inaccuracy of the formulae due to Maxwell and Palmer for the strip-above-ground system are confirmed. A new and simple method is derived for calculating approximate values for the pseudo-TEM impedance, wavelength and attenuation of a non-homogeneous dielectric-slab-supported strip-above-ground system (e.g. Microstrip). The numerical results obtained by this method, which depends only on a knowledge of the values for air-spaced lines, are in good agreement with measured results. As a result, it appears that the dominant mode propagating in a Microstrip line is more closely TEM in character than was hitherto supposed. In theory, however, there must exist a longitudinal component of the electromagnetic field if the boundary conditions are to be satisfied. Measurements on the electrolytic tank of coupling between adjacent conductors and of field spread are also described. From the analysis as a whole it would appear that the tri-plate (sandwich line) has certain theoretical advantages over the strip-above-ground system. However, these are likely to be offset by a number of important practical disadvantages.

(1) LIST OF PRINCIPAL SYMBOLS

(1.1) Strip Transmission Lines

E_n = Voltage gradient at a metallic boundary (assumed normal to the boundary).

H_t = Magnetic intensity at a metallic boundary (assumed tangential to the boundary).

I_l = R.F. current flowing on surface of boundary (assumed longitudinal in direction of propagation).

V = R.F. potential between two strip conductors.

ϵ_0 = Permittivity of free space, F/m.

ϵ_d = Permittivity of a specified dielectric medium other than free space, F/m.

μ_0 = Magnetic permeability of free space, H/m.

μ_m = Magnetic permeability of metallic boundary, H/m.

σ_m = Conductivity of metallic boundary, mhos/cm.

f = Frequency, c/s.

c = Velocity of light in free space.

v = Phase velocity of transmission line.

$Z_\eta = (\mu_d/\epsilon_d)^{1/2}$ = Wave impedance in the dielectric medium, ohms.

Z_0 = Characteristic impedance of transmission line.

Z_0' = Characteristic impedance of transmission line of Fig. 4.

$R_s = (\pi f \mu_m/\sigma_m)^{1/2}$ = R.F. skin resistance of conductor, ohms.

α_m = Attenuation coefficient of metallic losses.

α_d = Attenuation coefficient of dielectric losses.

α_m' and α_d' = Attenuation coefficient for the transmission line of Fig. 4.

a, b, h and t = Physical dimensions of a strip transmission line (see Fig. 1).

s = A distance measured along the surface of a conductor in the plane of the cross-section.

M, N = Constants which are a function of the geometry of the transmission line [see eqns. (4) and (5)].

(1.2) Electrolytic Tank Analogue

E_n = Voltage gradient in electrolyte at the surface of an electrode (assumed normal to the surface).

I = Current flowing from metallic electrode into surrounding electrolyte.

V = Voltage between two specified electrodes.

ρ = Resistivity of the electrolyte, ohm-cm.

s = A distance measured along the surface of an electrode in the plane of the electrolyte.

S_r = Surface perimeter of the rth electrode in the plane of the electrolyte.

M, N = Constants which are a function of the geometry of the electrode system [see eqns. (9) and (10)].

k_r = Form factor for the rth electrode [see eqn. (22)].

(2) INTRODUCTION

The high cost, bulk and ever-increasing complexity of equipment fabricated in standard waveguide has stimulated interest in several alternative types of transmission line. One of these is the so-called planar or strip transmission line,[1–12] several alternative forms of which are illustrated in Fig. 1. Counterparts of most coaxial or waveguide components, such as hybrid junctions, directional couplers, etc., can be realized in planar form, and this has suggested the possibility of fabricating quite complicated component assemblies at relatively low cost by the use of conventional printed-circuit techniques.[13,14] These component assemblies would, moreover, have the advantage of relatively small bulk and weight.

Despite the physical simplicity of these transmission lines the rigorous mathematical analysis of their properties presents considerable difficulties. When the strip conductors are embedded in a homogeneous isotropic dielectric medium of infinite extent it is possible to assume uniform propagation in the TEM-mode provided that certain critical dimensions do not exceed an appreciable fraction of a wavelength. This reduces the problem to a 2-dimensional analysis of the static field between parallel plates. Even so, owing to the discontinuous character of the metallic boundaries, considerable difficulty has been experienced

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Mr. Dukes is with Standard Telecommunication Laboratories.

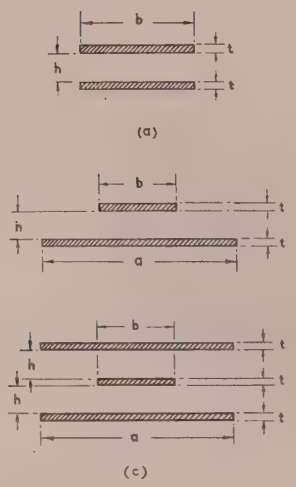


Fig. 1.—Cross-sections of transmission lines studied.

(*a*) Balanced parallel-plate line. Measurements were limited to the range $h/b = 0{\cdot}05$ to $h/b = 0{\cdot}8$ for $b/t = 32$, 64 and 96.
(*b*) Strip above finite ground. Measurements were limited to the range $h/b = 0{\cdot}15$ to $h/b = 1{\cdot}0$ for $a/b = 2$ and $a/b = 3$, with $b/t = 32$ throughout.
(*c*) Sandwich or tri-plate line. The outer conductors are at the same potential. Measurements were limited to the range $h/b = 0{\cdot}125$ to $h/b = 1{\cdot}0$ for $a/b = 1$, 2 and 3, with $b/t = 32$ throughout.

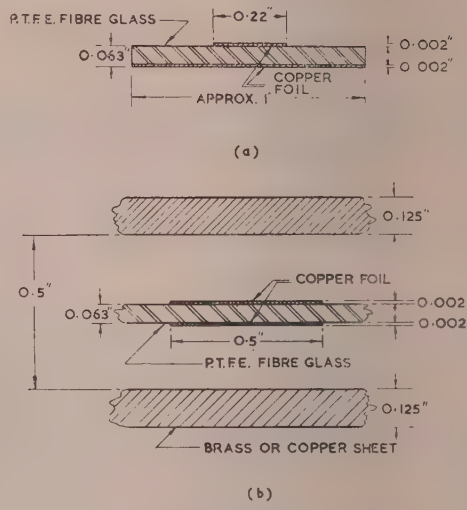


Fig. 2.—Examples of practical strip transmission lines with typical dimensions.

(*a*) Microstrip transmission line.
(*b*) Tri-plate, sandwich or shielded-strip transmission line.

in obtaining solutions capable of universal application. Solutions of differing rigour have been obtained for certain areas of restricted interest, but in many practical instances the solutions are of questionable accuracy. For example, the impedance range of 50–150 ohms is usually outside the nominal range of validity, and secondly, for the lines shown in Figs. 1(*a*) and 1(*c*) the formulae generally exclude values of a other than $a = \infty$.

In view of this uncertainty it was decided to analyse the properties of the lines shown in Fig. 1 with the aid of an electrolytic tank. The method provides an exact analogue of the electric field in any cross-section of the line, and may therefore be used to evaluate numerical results whose accuracy depends only on the degree of precision employed. In the present instance, because of the wide field studied an intermediate standard of accuracy was considered sufficient. This allowed the development of a new technique which avoided the use of movable probes and permitted a very rapid evaluation of quantities such as attenuation.

Unfortunately the practical application of all solutions based on the TEM approach is somewhat questionable. This is so because practical strip transmission lines such as those illustrated in Fig. 2 involve—to varying extents—dielectric supporting slabs having values of ϵ_d different from that of the surrounding space. The presence of a longitudinal dielectric interface necessitates, as a minimum, the existence of a longitudinal component of the electric vector[7,8,15] (for a demonstration of this proposition, see Appendix 13.1). In this case a TEM solution is theoretically inadmissible, and the nature of the dominant mode must be determined by direct application of the Maxwell equations for the particular boundary conditions chosen.

Despite this fundamental limitation the results contained in the paper are of considerable interest for several reasons. First, a rigorous modal solution as defined above has only been achieved in one special case,[7,8] and because of the many simplifying assumptions involved, the solution is of questionable application. Secondly, failing a modal solution of universal application the TEM approach provides the only remaining basis for design. Thirdly, it is important to determine to what extent a disagreement between measured and computed quantities is due to a general inapplicability of the TEM approach and to what extent it may be due to insufficient rigour of the particular TEM formulation employed. The analysis of Section 8.1 suggests the latter to be the case.

(3) OUTLINE OF THE PROPERTIES OF STRIP TRANSMISSION LINES

A convenient starting-point in any discussion of strip transmission lines is the infinite parallel-plate line. Both plates are of infinite extent in the positive and negative x-directions, and the wave is assumed to propagate uniformly in the z-direction. In addition it is assumed that the spacing between the plates is small compared to the wavelength. The transmission line has zero characteristic impedance, of course, and for finite values of E and H the power transmitted is infinite.

A first-order approximation to the parallel-plate transmission line of Fig. 1(*a*) can be achieved by taking a segment of the infinite parallel-plate transmission line and lining each exposed side of the segment with a hypothetical magnetic boundary. This is illustrated in Fig. 3. By this artifice the magnitude and orientation of both the electric and magnetic vectors remains unchanged. The solution[16] for this hypothetical line is

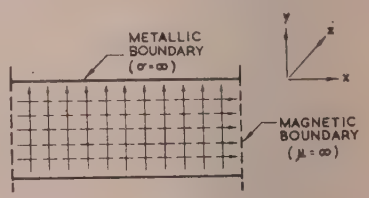


Fig. 3.—Hypothetical parallel-plate transmission line used as a reference standard.

↑ Lines of electric force E.
→ Lines of magnetic force H.
The wave is shown propagating in the z-direction.

$$E = Z_\eta H \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$Z_0 = Z_\eta h/b \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2a)$$

$$= V/I_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2b)$$

$$\alpha_m = R_s(Z_\eta h)^{-1} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where the symbols have the meaning outlined in Section 1.1. The time-variant terms of the form $\varepsilon^{-j2\pi ft}$ are omitted for simplicity, and the symbols E_r, H_t, V and I_e are assumed throughout to represent the r.m.s. magnitudes of the quantities concerned.

This hypothetical line represents a convenient reference standard whose properties are easily calculated. For example, for an air-spaced line with copper conductors of dimensions $b = h = 1$ cm operated at a frequency of 1 000 Mc/s, we have

$$\sigma_m = 5{\cdot}8 \times 10^7 \text{ mhos/cm}$$
$$\epsilon_0 = 8{\cdot}85 \times 10^{-12} \text{ F/m}$$
$$\mu_m = \mu_0 = 1{\cdot}257 \times 10^{-6} \text{ H/m}$$
$$Z_0 = Z_\eta = 377 \text{ ohms}$$
$$R_s = 8{\cdot}24 \times 10^{-3} \text{ ohm}$$
$$\alpha_m = 2{\cdot}18 \times 10^{-3} \text{ neper/m} = 1{\cdot}89 \times 10^{-2} \text{ dB/m}$$

In the work that follows the attenuation of any given configuration will be referred to that of the hypothetical line having the same values of h and b, and constructed of the same materials. The impedance and attenuation of the reference line will be denoted by the symbols Z_0' and α_m'.

If the side walls of this hypothetical line are now removed, and if for the sake of generality the size of one conductor is altered, the field takes up the approximate form shown in Fig. 4.

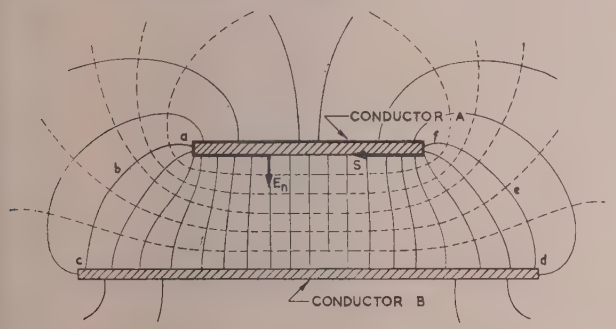


Fig. 4.—Sketch of field configuration for a typical strip transmission line.

E_n is the voltage gradient at the surface of the conductor (assumed normal to the conductor). The heavy line indicates the closed integration path s around the surface of the conductor.

It will be evident that some of the r.f. current must flow on the outer surfaces of the strip conductors.

To calculate the impedance and attenuation of such a line use is made of the following relationships, the detailed derivations of which are given in Appendix 13.1:

$$Z_0 = Z_\eta M \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

$$\alpha_m = R_s N / Z_\eta \text{ nepers/metre} \quad . \quad . \quad . \quad . \quad (5)$$

where M is a dimensionless constant depending only on the geometry of the system and is given by

$$M = V \Big/ \oint_A E_n ds = V \Big/ \oint_B E_n ds \quad . \quad . \quad . \quad (6)$$

and N is a quantity with the dimension of the reciprocal of distance, and depends partly on the geometry of the system and partly on its size as follows:

$$N = \tfrac{1}{2}\Big(\oint_A E_n^2 ds + \oint_B E_n^2 ds\Big) \Big/ \Big(V \oint_A E_n ds\Big). \quad . \quad (7)$$

In the above equations E_n is the electric gradient at the surface of the conductor (assumed normal to the surface) and the subscripts A and B denote the appropriate conductors in Fig. 4. The integrals are taken round the closed path indicated by the heavy line for conductor A. It is assumed that the two conductors are unique, and hence

$$\oint_A E_n ds = \oint_B E_n ds$$

In the case of the line illustrated in Fig. 1(*c*), eqns. (6) and (7) must be modified to allow for the additional conductor. The nature of this change will be apparent on considering the effect of arbitrarily dividing, say, conductor A of Fig. 4 into two.

(4) ANALOGY OF THE ELECTROLYTIC TANK

The electrolytic tank, constructional details of which are given below, provides a thin resistive layer which, ideally, is of infinite extent. If two electrodes of opposite polarity are placed in contact with this layer, currents will flow between the electrodes through the resistive layer. Thus Fig. 4 may accordingly be regarded as a flux plot for two rectangular electrodes in contact with a resistive layer lying in the plane of the paper. The lines of voltage gradient E are also the lines of current flow, and the dashed lines represent equipotential lines.

It is an essential assumption of TEM propagation that in any cross-section the Laplace conditions

$$\frac{\partial^2 E_x}{\partial x^2} + \frac{\partial^2 E_x}{\partial y^2} = \frac{\partial^2 E_y}{\partial x^2} + \frac{\partial^2 E_y}{\partial y^2} = 0 \quad . \quad . \quad . \quad (8)$$

must be satisfied at each and every point throughout that cross-section. In the case of the resistive layer it is also necessary for the field to be Laplacian. From this it follows that, given the same boundary conditions, the field plots must be identical for the two cases. Thus from an analogue on the electrolytic tank we can obtain a plot of the field conditions existing within a strip transmission line.

In fact, if we wish to obtain only the impedance and the attenuation of the line, it is not necessary to plot the field in its entirety. As can be seen from eqns. (6) and (7), all that is required is a knowledge of the conditions in the immediate proximity of the conductor surfaces. In Appendix 13.1 it is shown that, for a resistive layer of resistivity ρ, the constants M and N assume the following form:

$$M = V \Big/ \oint_A E_n ds = R/\rho \quad . \quad . \quad . \quad . \quad (9)$$

where $R = V/I$ is the measured resistance between the electrodes, and

$$N = \tfrac{1}{2}\Big(\oint_A E_n^2 ds + \oint_B E_n^2 ds\Big) \Big/ \Big(V \oint_A E_n ds\Big) \quad . \quad (10a)$$

$$= \frac{\rho}{2VI}\left[\sum_A (\delta I)^2/\delta s + \sum_B (\delta I)^2/\delta s\right] \quad . \quad . \quad (10b)$$

In the second form of eqn. (10) it is assumed that the surface of the electrode is broken up into a number of elementary segments as illustrated in Fig. 6, each of constant width δs. The current flowing out from each segment into the surrounding resistive sheet is δI.

For reasons which are discussed below it has been found preferable to measure the electrode elementary currents δI rather than the voltage gradient E_n. Hence eqn. (10*b*) provides in the present instance the more useful formulation.

(5) THE ELECTROLYTIC TANK

(5.1) Construction of the Tank

Apart from certain minor mechanical differences, the tank, which is illustrated in Fig. 5, was identical to that designed by Cherry, Boothroyd and Makar[18] for use in network analysis and

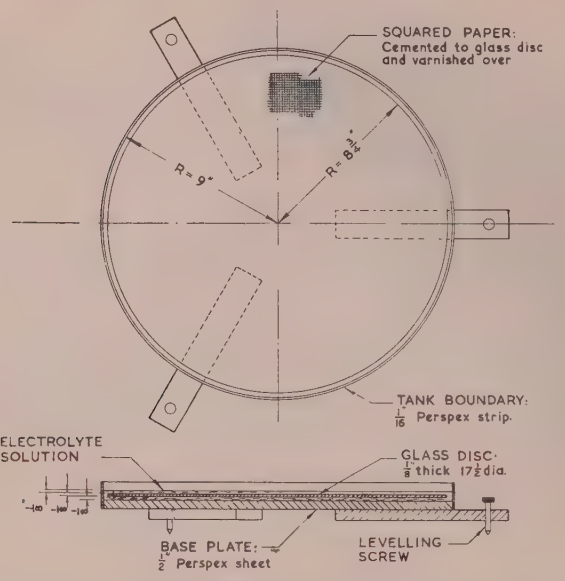


Fig. 5.—Construction of the electrolytic tank.

The glass disc is supported by three small spacers each $\frac{1}{8}$ in thick. Details of the overhead system of cross-bars for positioning the electrodes in the tank are omitted.

synthesis. It consists essentially of two thin circular layers of fluid electrolyte (copper sulphate) connected together at their peripheries. By this means it is possible to obtain a very good approximation to an infinite medium. For example, if current flows between one electrode at the centre of the upper layer and one electrode at the centre of the lower layer, the error in the potential of the field at the edge of the tank is reported[18] to be less than 1%. (In the present instance, of course, the electrodes are all situated in the upper layer.)

Certain mathematical transformations[19] are possible which transform one half of the infinite plane into a finite bounded region. This is convenient in that it avoids any need for a double layer of electrolyte. On the other hand, a transformed representation would complicate the design of the electrodes considerably, since in this case they would be no longer planar, nor would they be of constant thickness.

(5.2) Design of the Electrodes

Fig. 6 illustrates the design of the planar electrodes which are inserted into the upper layer of the tank so as to provide a representation of the cross-section of the Microstrip line. The formation of the copper segments was accomplished by one of the well-known printed-circuit processes. The choice of suitable dimensions was regulated by three principal considerations. First, the width should be such that a convenient range of lines as in Fig. 1 could be covered using combinations of one standard electrode. Secondly, the width should be such that a variety of voltage standing-wave ratios could be accommodated satisfactorily on the tank. Thirdly, the thickness should be of the same order as might be encountered in a practical strip transmission line (relative to the width). The choice of segment size

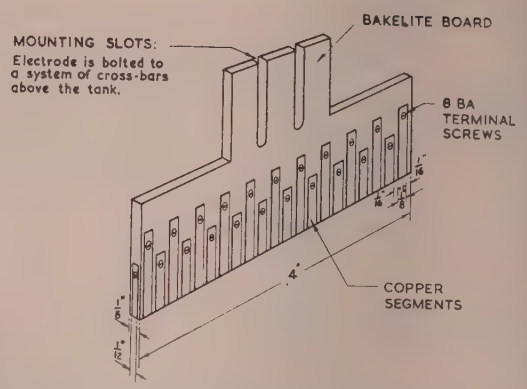


Fig. 6.—Sketch of typical segmented electrode.

The copper segments were deposited by a printed-circuit technique.

and spacing was largely a matter of chance; there happened to be available at that time a silk screen designed to print segments of this particular size.

In practice it was found necessary to construct only three of these segmented electrodes. When building up a configuration of the type shown in Fig. 1(*c*), the symmetry was such as to permit the extensive use of solid metallic electrodes. To support the electrodes a grid of movable horizontal bars was provided. These were hinged at one side so that the assembly of electrodes could be lifted clear of the tank without disturbance to the actual arrangement.

The decision to measure the currents flowing in and out of the electrodes rather than the voltage gradient at the surface was influenced by the following considerations. First, in order to measure the surface voltage gradient, two very fine closely-spaced electrodes are required, which must be positioned accurately in the immediate vicinity of the electrode, and orientated carefully at right angles to the electrode surface. (One single probe may be used and the voltage measured between it and the electrode. This, however, introduces certain supplementary difficulties.) Secondly, since it was planned to cover a large number of possible configurations, a rapid method of measurement was essential. The voltage-probe method would necessitate an overhead system for locating the probe quickly and positively in different positions. To cover all the configurations envisaged, it seemed that this device would be rather complex and would involve a certain amount of specialized mechanical design. Finally, although a double-probe method is ultimately the most accurate—since it is dependent only on the field in the electrolyte and is not affected by polarization potentials between the electrodes and the electrolyte—preliminary tests showed that an accuracy of the order sought for could be attained by the method actually employed.

(5.3) Electrical Circuit and Switching System

The electrical schematic is shown in Fig. 7. The terminal points on the right-hand side of the switching array are connected to the electrode segments, and on the other side they are connected in parallel to the supply lines. The ammeter is inserted in place of each short-circuiting link in turn, thus permitting successive measurement of the current into, or out of, each segment of the electrodes.

(5.4) Sources of Error

There are a number of possible sources of error in this type of measurement, which are discussed in Appendix 13.2.

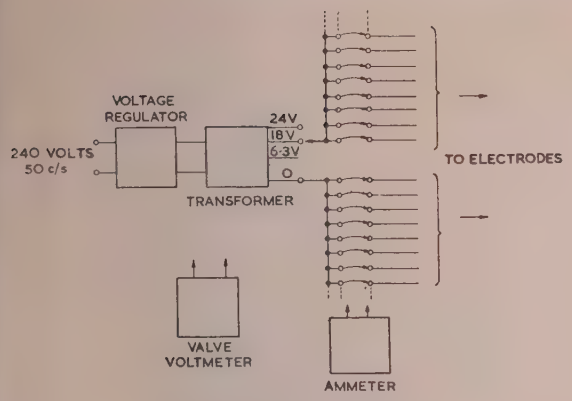


Fig. 7.—Electrical schematic.

(5.5) Operation of the Tank

The operating routine is fairly straightforward. The tank and the electrodes must be kept perfectly clean, and the electrolyte solution must be thoroughly mixed before use. The tank must be carefully levelled by means of the screws provided, and the thickness of the two layers of electrolyte must be accurately maintained.

To calibrate the tank, two electrodes representing a configuration for which the field solution is known are inserted in the tank, and the voltage across and the current between the electrodes are measured. Care should be taken that the surface current density is a good deal less than the permissible maximum, namely 0·3 mA/mm². The most convenient configuration is two identical cylinders (about $\frac{3}{4}$ in diameter) placed anywhere from 4 to 12 in apart. The solution for this is

$$\rho = \pi R / [\text{arc cosh}\,(h/2r)] \quad . \quad . \quad . \quad . \quad (11)$$

where $R = V/I$ is the measured resistance between the electrodes, h is the spacing between them, r is their radius, and ρ is the resistivity of the electrolyte sheet (the resistance between two sides of a square of any size).

(6) MEASUREMENTS OF CHARACTERISTIC IMPEDANCE

Eqns. (4) and (9) show that the characteristic impedance may be determined from a simple measurement of the total current flowing between the electrodes. No difficulty is presented by these measurements, which may be performed very rapidly. The results are presented in Figs. 8–10, and compared with various solutions of differing rigour developed by different workers.

The upper pair of curves for Fig. 8 are plotted from expressions quoted by Assadourian and Rimai[2] for parallel plates of zero thickness. The first is due to Maxwell, and is as follows:

$$Z_0/Z_0' = 2\Big/\{1 + h/(\pi b)[1 + \log\,(1 + \pi b/h)]\} \quad . \quad (12)$$

where Z_0' is the characteristic impedance of the hypothetical line of Fig. 3 and eqns. (1)–(3). The second curve is a plot of an expression due to Palmer:

$$Z_0/Z_0' = 2\Big/\{1 + h/(\pi b)[1 + \log\,(2\pi b/h)]\} \quad . \quad (13)$$

However, both these solutions are only valid for $h/b \ll 1$. For two narrow strips Assadourian and Rimai[2] give the following expression, also plotted in Fig. 8:

$$Z_0/Z_0' = (2b/\pi h)\log\,(4h/b) \quad . \quad . \quad . \quad (14)$$

As is to be expected, the measured curve tends to pass from the wide- to the narrow-strip solution as the spacing is increased.

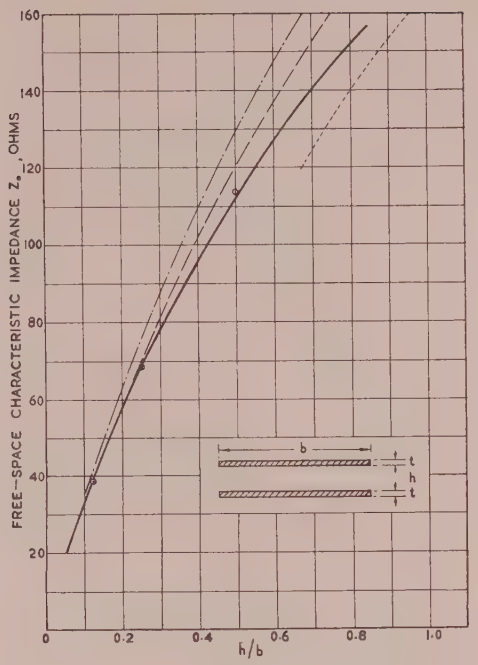


Fig. 8.—Impedance of air-spaced parallel-plate line.
——— Measured on electrolytic tank for b/t = 32, 64 or 96.
—·— Maxwell's solution.
— — Palmer's solution.
- - - - Narrow strip above ground.
○ Black and Higgins.

Recently a new solution has been presented by Black and Higgins.[20] Their method depends on a rather more involved application of the Schwartz–Christoffel transformation and leads to results which are inherently more accurate. Some numerical results from their paper are plotted in Fig. 8. As will be seen, the points lie quite close to the measured curve. In view of the estimated accuracy of measurement (±3% for impedance) the discrepancy is not significant.

Using the method of images the above expressions can be transformed to provide solutions for a strip conductor above a ground plane of infinite width. Eqns. (12)–(14) accordingly become

$$Z_0/Z_0' = 1\Big/\{1 + 2h/(\pi b)[1 + \log\,(1 + \pi b/2h]\} \quad (15)$$

$$Z_0/Z_0' = 1\Big/\{1 + 2h/(\pi b)[1 + \log\,(\pi b/h)]\} \quad . \quad . \quad (16)$$

$$Z_0/Z_0' = (b/2\pi h)\log\,(8h/b) \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

These three solutions are plotted in Fig. 9 together with measured results for two cases of a strip conductor above a ground plane of finite width. Here the discrepancy is even larger; first, eqns. (15) and (16) are only valid for $h/b \ll 1$, and secondly, they only apply when the width of the ground plane is much greater than that of the strip conductor.

Black and Higgins[20] have given some numerical results for lines having ground planes of finite width. These results are plotted in Fig. 9, and once again they are in striking agreement with measurements made on the electrolytic tank. It should be noted that Black and Higgins also calculated the lower bound for $a/b = \infty$, and that this lower bound tends to lie very close to the measured curve for $a/b = 3$. The fact that their results for $a/b = 4$ and 8 tend to fall *below* their lower bound is due to certain approximations adopted by them in order to allow

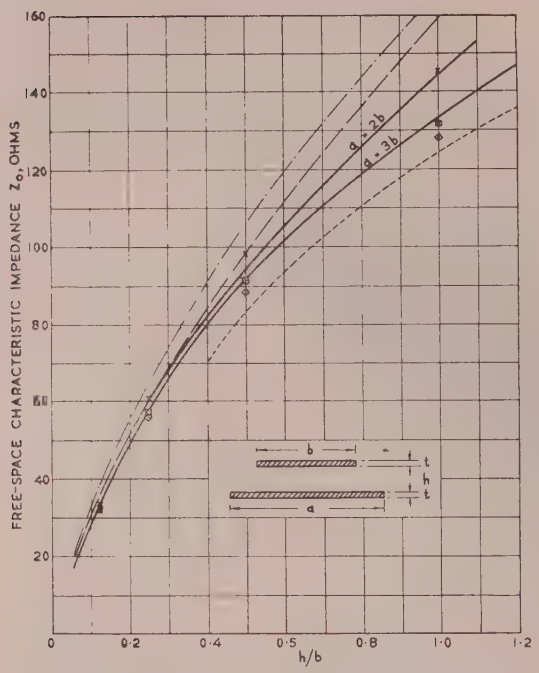


Fig. 9.—Impedance of air-spaced strip above ground.

Measured on electrolytic tank for $b/t = 32$.
Maxwell's solution.
Palmer's solution.
Narrow strip above ground.
Black and Higgins, $a = 2b$.
Black and Higgins, $a = 4b$.
Black and Higgins, $a = 8b$.

numerical computation with the aid of a computing machine of standard design.

Fig. 10 relates to the sandwich line of Fig. 1(*c*). The results are compared with a few numerical values selected from curves given by Cohen[21,22] for a strip of small but finite thickness between two finite parallel plates. When constructing these curves, Cohen had available several solutions derived independently by E. Oberhettinger and W. Magnus, N. A. Begovich, R. L. Pease and A. A. Oliner. For the relatively simple case of a strip conductor of zero thickness the following exact formula is valid:

$$Z_0 = \frac{30\pi \mathrm{K}(k)}{\mathrm{K}(k')} \quad . \quad . \quad . \quad . \quad . \quad (18)$$

where K(k) and K(k') are complete elliptic integrals of the first kind, and

$$k = \operatorname{sech} \pi b/4h \quad . \quad . \quad . \quad . \quad . \quad (19)$$

$$k' = \tanh \pi b/4h \quad . \quad . \quad . \quad . \quad . \quad (20)$$

The close correspondence between the measured results for $a = 2b$ and $a = 3b$ with the theoretical values for infinite ground planes is not altogether surprising. During the experiments described in Section 7.2 it was found that the current distribution on the surface of the strip conductor varied considerably less with changes in ground-plane width for the sandwich line of Fig. 1(*c*) than for the strip-above-ground line of Fig. 1(*b*).

(7) ATTENUATION MEASUREMENTS

(7.1) Preliminary Observations

To obtain the attenuation of a given line, the constant N of eqn. 10(*b*) is first determined and is then substituted in eqn. (5).

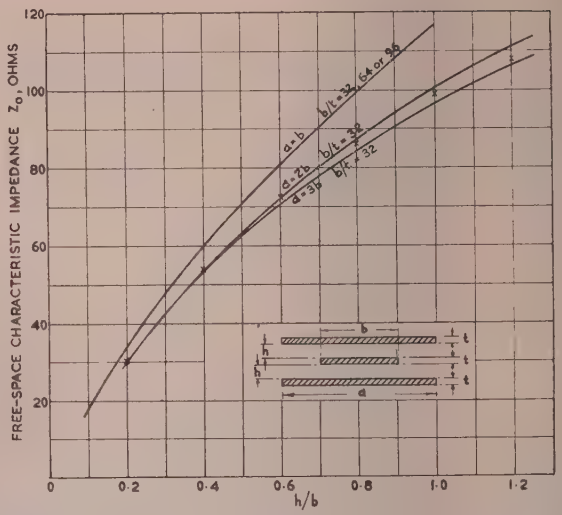


Fig. 10.—Impedance of tri-plate (or sandwich) transmission line.

Measured on electrolytic tank.
× Representative points from Cohen's curves.

The result, however, is a function of both frequency and linear size; it is convenient, therefore, to normalize with respect to the hypothetical line of Fig. 3, by dividing eqn. (5) by eqn. (3):

$$\alpha_m/\alpha'_m = hN \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

At the same time it is convenient to re-express N in a general form useful for a transmission-line system comprising any number of "go" and "return" conductors. First, we define a form factor k_r for the rth electrode (representing the rth conductor) as follows:

$$k_r = n_r \sum_r (\delta I)^2 \Big/ \Big(\sum_r \delta I\Big)^2 \quad . \quad . \quad . \quad . \quad (22)$$

where n_r is the number of segments on the rth electrode, and $\sum_r$ indicates summation round the cross-sectional perimeter of the rth electode (see Fig. 4). The segments are assumed to be of constant width. The form factor k_r will be recognized as the ratio of the mean-square current to the square of the mean current. As such it provides a direct measure of the copper loss for the conductor in question relative to an otherwise identical conductor for which the current distribution is uniform over its surface.

Summing over all the m conductors in the system, the following expression for N is obtained:

$$N = \tfrac{1}{2}(\rho/R) \sum_m (q_r k_r/S_r) \quad . \quad . \quad . \quad . \quad (23)$$

where q_r is the fractional current carried by the rth electrode and S_r is the surface perimeter of the rth electrode in cross-section.

(If one or two of the segments are slightly smaller than the remainder—as occurs in Fig. 6—the $(\delta I)^2$ contributions of the segments in question may be multiplied by $(\delta S)/(\delta S)'$, where $(\delta S)'$ is the width of the smaller segments. This method of correction, which involves several approximations, is quite accurate provided that the current contributions of the segments in question are only a small fraction of the total carried by the electrode.)

(7.2) Typical Current Distributions

In Fig. 11 are plotted three typical current distributions for one-half of one conductor of the parallel-plate line of Fig. 1(*a*).

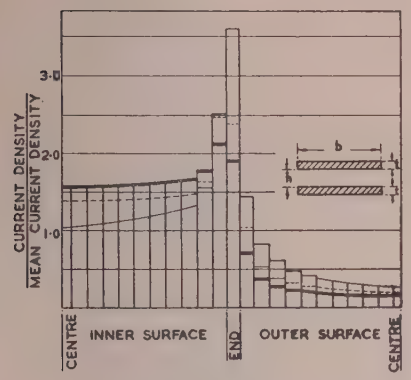


Fig. 11.—Conductor current distribution for parallel-plate transmission line of Fig. 1(*a*).

—— $h/b = 0{\cdot}70$ ($k_r = 1{\cdot}49$).
- - - - $h/b = 0{\cdot}30$ ($k_r = 1{\cdot}51$).
▬▬ $h/b = 0{\cdot}167$ ($k_r = 1{\cdot}58$).
($b/t = 32$ throughout).

It should be stressed that the ordinate represents currents normalized with respect to the mean current density. It is interesting to note the manner in which the distribution varies for different ratios of spacing to width. For wide spacings, e.g. $h/b \to 1$, a very much larger proportion of the current is carried by the end and outer surfaces. A similar tendency is observed in the case of the strip-above-ground system of Fig. 1(*b*) when the ground-plane width is increased in steps from $a = b$ to $a = 3b$.

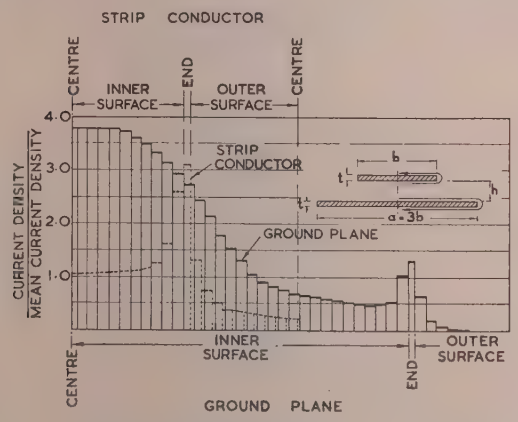


Fig. 12.—Conductor current distribution for strip above ground with $a = 3b$, $h/b = 0{\cdot}3$ and $b/t = 32$.

- - - - Strip conductor ($k_r = 1{\cdot}50$).
▬▬ Ground plane ($k_r = 2{\cdot}76$).

Fig. 12 shows the plots for both conductors of the strip-above-ground system of Fig. 1(*b*). The results given by this series of measurements have also been used to compute the radiation and coupling through slots in the ground plane of a practical Microstrip line as shown[23] in Fig. 2(*b*). Despite the approximation involved in assuming a homogeneous dielectric medium of infinite extent, the computed values of coupling are in good agreement with measurements at microwave frequencies.

As will be evident from Figs. 11 and 12, the segmentation is so coarse that a considerable sampling error is likely to occur owing to loss of detail. The magnitude of this error was

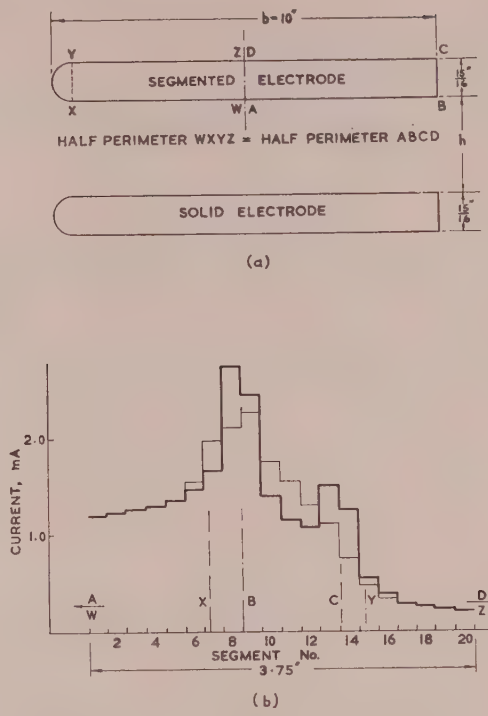


Fig. 13.—Detailed investigation of end currents.

(*a*) Large-scale analogue of parallel-plate line. Half-perimeter WXYZ = Half-perimeter ABCD.
(*b*) Distribution of end currents for $h/b = 0{\cdot}4$.
▬▬ Square end piece.
—— Semicircular end piece.

investigated with the aid of the large-scale analogue of Fig. 13(*a*), which was equipped with both semicircular and square end-pieces. Typical distribution curves for the current in the end segments are illustrated in Fig. 13(*b*).

Five sets of readings were taken for h/b ratios ranging from 0·06 to 0·6. The readings for the square-ended half-perimeter were analysed in two ways. First, the form factor was calculated from the readings as they stood. A modified form factor was then computed by taking the segment currents consolidated in groups of five at a time, thus simulating a coarsely-segmented electrode. On examination it was found that the form-factor sampling error was a function principally of the ratio of the end-segment current to the total electrode current. This gave

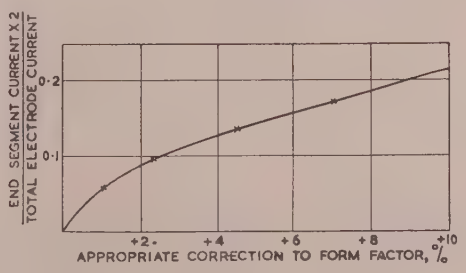


Fig. 14.—Form-factor correction curve.

the correction curve of Fig. 14, which was used in the subsequent calculations of attenuation.

The effect of edge shape on form factor was found to be small.

Over the above range the form factor of the half line equipped with semicircular end pieces was consistently about 3% less than that for the square-ended half line.

In using these current-density plots it may be assumed that the current at the centre of the inner surface is that which would occur if the conductors were of infinite width. This assumption introduces negligible error for the h/b range studied.

(7.3) Attenuation Results

The results of these experiments, illustrated in Figs. 15–17, are largely self-explanatory. The technique of measurement and computation is that described in Sections 7.1 and 7.2, and in all cases the results have been corrected for end-effect sampling error.

The first line studied was the parallel-plate type (Fig. 15). In order to estimate the accuracy of the technique rather a large

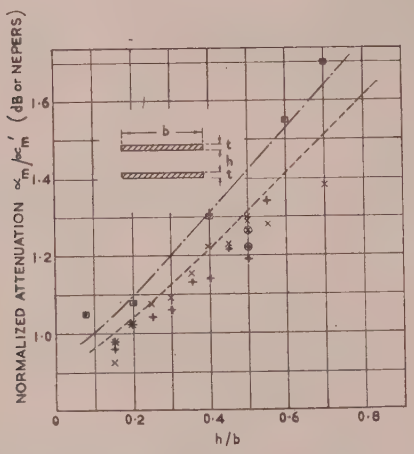


Fig. 15.—Relative attenuation of parallel-plate line (metallic losses only).

---- Assadourian and Rimai, for $b/t = 32$.
\+ Measured results (upper electrode), $b/t = 32$.
× Measured results (lower electrode), $b/t = 32$.
⊕ ⊗ } Special readings taken with ammeter drop balanced out.
—·— Assadourian and Rimai, for $b/t = 64$.
⊡ Measured results, $b/t = 64$.

number of readings were taken. As an additional precaution, readings were taken with two different but nominally identical electrodes. The possibility of a significant error due to ammeter voltage drop (Appendix 13.2) was also examined. However, as can be seen from Fig. 15, the spread of the points was largely random in character. The principal cause of this randomness was the rapidity with which the electrodes corroded if left in the tank for any length of time. In subsequent experiments they were cleaned more often, and it is noticeable that the later readings tend to show less spread. These and other considerations outlined in Appendix 13.2 suggest that the attenuation results are accurate within outside limits of $\pm 8\%$.

Also plotted in Fig. 15 are curves computed from formulae given by Assadourian and Rimai[2] for a wide strip above ground. These formulae, quoted below, may be adapted without difficulty (by the method of images) to the case of two parallel plates of equal width. It will be noted that the agreement here is, on the whole, closer than for the characteristic-impedance curves of Fig. 8.

The results for a strip above ground are illustrated in Fig. 16. Here the agreement is not so good. This may be due in part to

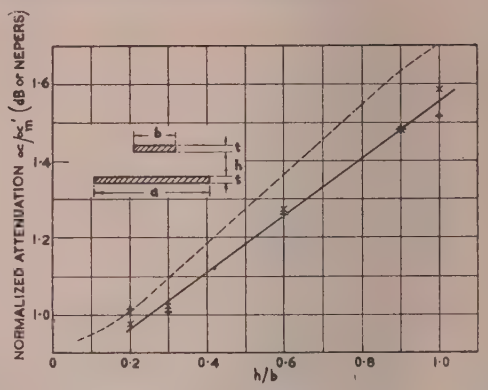


Fig. 16.—Relative attenuation of strip-above-ground system (metallic losses only).

---- Assadourian and Rimai, for $b/t = 32$ and $a = \infty$.
× Measured results: $b/t = 32$, $a = 2b$.
\+ Measured results: $b/t = 32$, $a = 3b$.
—— Mean straight line through measured points.

the finite width of the ground plane. The relevant formulae are as follows:

$$\alpha_{mg}/\alpha_m' = (1 + \pi b/2h)/[1 + \pi b/2h + \log(1 + \pi b/2h)] \quad . \quad . \quad . \quad . \quad (24a)$$

$$\alpha_{ms}/\alpha_m' = (1 + \pi b/2h + \pi - 2\log \delta/4)/[1 + \pi b/2h + \log(1 + \pi b/2h)] \quad (24b)$$

where α_{mg} = Attenuation due to ground plane losses.

α_{ms} = Attenuation due to strip conductor losses.

$(\alpha_{mg} + \alpha_{ms})/\alpha_m'$ = Total attenuation of transmission line relative to hypothetical line of Fig. 4.

$$\delta = p - 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

$$p = -1 + 2\beta^2 + 2\beta(\beta^2 - 1)^{1/2} \quad . \quad . \quad . \quad (26)$$

$$\beta = 1 + t/h \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

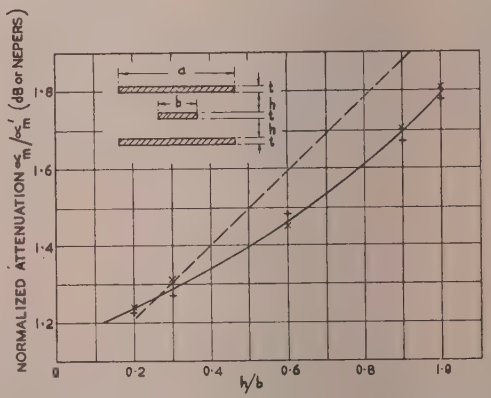


Fig. 17.—Relative attenuation of tri-plate or sandwich line (metallic losses only).

---- From curves by Cohen.
× Measured results: $b/t = 32$, $a = 2b$.
\+ Measured results: $b/t = 32$, $a = 3b$.
—— Mean curve through measured points.

Some results for the tri-plate line are presented in Fig. 17. Unfortunately there does not as yet exist a rigorous solution for the range of h/b values studied. Cohen[22] has produced a set of curves interpolated between two solutions, one for a wide

strip and the other for a narrow strip. For the wide strip, which is the case of major interest, Cohen[22] gives the following formula:

$$\alpha_m = \frac{2R_s Z_0}{Z_\eta^2 h}\left\{1 + b/h + \frac{1}{\pi}(1 + t/h)\log(1 + 4h/t)\right\}$$

nepers/unit length . (28)

The transition region appears to centre on $h/b = 1{\cdot}4$, at which point the two solutions differ by about 10%. In this region Cohen[22] constructed what he considered to be a "reasonable" transition curve, but the accuracy of such a procedure is open to some doubt.

In conclusion it should be stressed that the method described here is inherently capable of any desired accuracy. However, in view of the wide field of interest and certain limitations in time and effort, the present accuracy was considered adequate.

(8) APPLICATION TO PRACTICAL COMPOSITE DIELECTRIC STRIP TRANSMISSION LINES

(8.1) Strip-above-Ground Systems

With one exception,[24] little interest has been shown in air-spaced strip-above-ground transmission lines, and hence no data are available. For the most part the form of construction used is that illustrated in Fig. 2(*a*). Such a line could be represented on the electrolytic tank by providing variations in thickness of the electrolyte corresponding to the regions of differing permittivity. By this means the capacitance C per unit length could be calculated, whilst the inductance L per unit length would be calculated for the free-space condition from the data of Figs. 8 and 9. The pseudo-TEM characteristic impedance and the pseudo-TEM velocity of propagation would then be given by

$$Z_0 = (L/C)^{1/2} \quad . \quad . \quad . \quad . \quad . \quad (29)$$

$$v = (LC)^{-1/2} \quad . \quad . \quad . \quad . \quad . \quad (30)$$

Unfortunately, as has already been emphasized in Section 2, this solution can only be assumed to be valid at very low frequencies. It was therefore decided that the questionable value of such data did not warrant proceeding with these measurements, which in view of the additional complications would have proved rather tedious. Moreover, it was realized that the data in hand were sufficient to provide upper- and lower-bound solutions, on the basis of which an intermediate pseudo-TEM solution of moderate rigour can be obtained, which as it happens is in rather good agreement with experimental results.

The first solution relates to the situation shown in Fig. 18(*a*). Since the dielectric medium is homogeneous throughout the entire cross-section, the line will therefore support a pure TEM-mode. The characteristic impedance for the air-spaced line is multiplied by $(\epsilon_0/\epsilon_d)^{1/2}$; the velocity of propagation is equal to that of light multiplied by the same factor; and the relative attenuation α_m/α'_m remains unaltered (provided that α'_m has been modified to allow for ϵ_d). This solution, for a given value of ϵ_d, provides a lower bound for the value of Z_0.

The upper bound solution is given by the hypothetical transmission line shown in Fig. 18(*b*). This effectively represents the limiting case where the relative permittivity ϵ_d/ϵ_0 is so large that the field outside the slab can be neglected. (For practical purposes it is sufficient that $\epsilon_d/\epsilon_0 \geqslant 10$.) The solution is derived rigorously from the solution of the tri-plate line of Fig. 1(*c*). The xz-plane through the centre of the strip conductor is a plane of symmetry, along which a magnetic boundary may be introduced without disturbance to the flux pattern. The upper half of the line may now be removed with the result shown

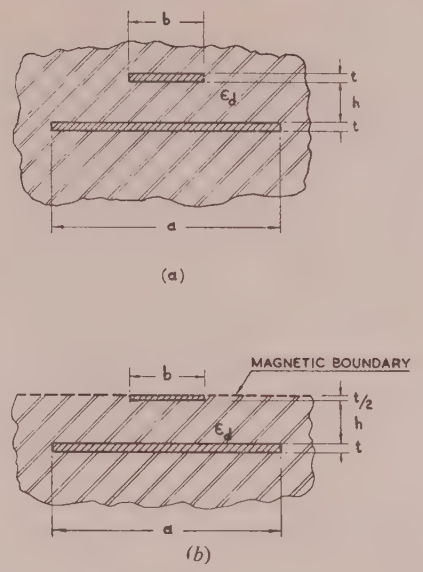


Fig. 18.—Composite dielectric line, *upper- and lower-bound* solutions.

(*a*) *Lower-bound solution.* The dielectric material is of infinite cross-sectional area.
(*b*) *Upper-bound solution.* This provides an exact picture of field conditions when $\epsilon_d/\epsilon_0 \to \infty$.

in Fig. 18(*b*). This line will likewise support a pure TEM-mode with the same velocity of propagation as for Fig. 18(*a*). The impedance is twice that of the corresponding tri-plate line multiplied by $(\epsilon_0/\epsilon_d)^{1/2}$, and the attenuation α_m/α'_m is equal to that of the tri-plate line.

It is shown in Appendix 13.1 that the capacitances of lines such as those shown in Fig. 19(*a*) and 19(*b*) can be calculated exactly

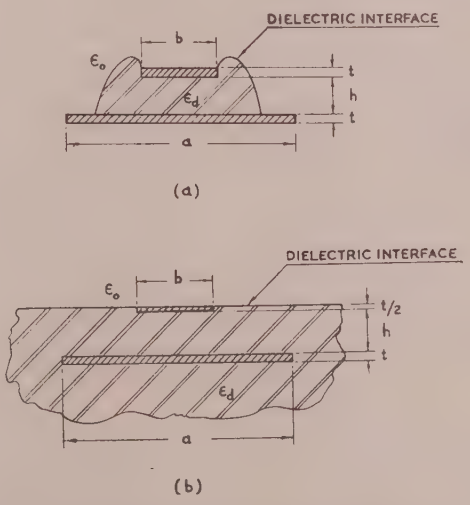


Fig. 19.—Composite dielectric lines—pseudo-TEM solution.

(*a*) Line with dielectric boundaries coincident with flux lines.
(*b*) Line for which pseudo-TEM solution is only approximate.

from a knowledge only of the appropriate upper and lower bound solutions, subject to one very important proviso. This proviso requires that the dielectric interface should lie along a flux line. From Fig. 4 it will be apparent that the dielectric support will have to assume a shape such as that shown in

Fig. 19(*a*). Application of the method to Fig. 19(*b*) can only yield a result which is approximately correct. Nevertheless it is believed that for practical dimensions and typical values of ϵ_d/ϵ_0 the capacitance of a practical Microstrip line as calculated by this method is in error from the true value by only a few per cent. This is substantiated by measurements of capacitance made on Microstrip lines generally, as in Fig. 2(*a*), but constructed from materials having relative permittivities between unity and 20. Some difficulty was experienced in eliminating stray capacitances, and proper allowance must be made for the fringing capacitance at each end of the line. Furthermore, for the measurements to be of value it is necessary to know the permittivities of the materials used to a high degree of accuracy. Although not wholly conclusive, these measurements suggest that the formulae given below provide characteristic impedances which are low by a maximum of 3% for $h/b = 0{\cdot}1$, increasing to 5% for $h/b = 1{\cdot}0$. These are outside figures, and there is good reason to believe that the error is in reality somewhat smaller, particularly for a small ϵ_d/ϵ_0 ratio.

Subject to the limitations in regard to the lie of the dielectric interface, the pseudo-TEM impedance $\bar{Z}_0$, and the pseudo-TEM velocity of propagation $\bar{v}$, are given rigorously by the following expressions:

$$\bar{Z}_0 = Z_0\xi \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (31)$$

where Z_0 is the characteristic impedance of the air-spaced strip-above-ground transmission line

and $$\bar{v} = c\xi \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (32)$$

where $$\xi = [1 + \nu^2(\epsilon_d/\epsilon_0 - 1)]^{-1/2} \quad . \quad . \quad . \quad (33)$$

and $$\nu = \frac{\text{Lower-bound impedance}}{\text{Upper-bound impedance}} \quad . \quad . \quad . \quad . \quad (34a)$$

For Fig. 19(*b*) the appropriate upper-bound solution is that of Fig. 18(*b*), and hence

$$v = \frac{Z_0, \text{ air spaced strip-above-ground}}{2Z_0, \text{ air spaced tri-plate line}} \quad . \quad . \quad (34)$$

Application of this method is illustrated for two types of fibre-glass dielectric in Figs. 20–22. Also shown are certain results due to Arditi,[4,5] who measured the capacitance C at low frequencies and the velocity of propagation v at microwave frequencies. On the assumption of TEM propagation it follows

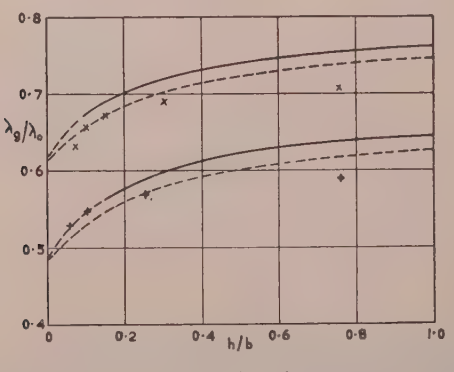


Fig. 20.—Comparison of measured and computed wavelength for composite dielectric Microstrip line (pseudo-TEM solution).

Upper curves, for $\epsilon_d/\epsilon_0 = 2{\cdot}65$.
Lower curves, for $\epsilon_d/\epsilon_0 = 4{\cdot}18$.
—— $a/b = \infty$, Black, Higgins and Cohen.
--- $a/b = 3$, electrolytic tank.
× Measured points for Teflon fibre glass with dimensions approximately as in Fig. 2(*a*) at 7300 Mc/s.
\+ Measured points for silicone fibre glass with dimensions approximately as in Fig. 2(*a*) at 4770 Mc/s.

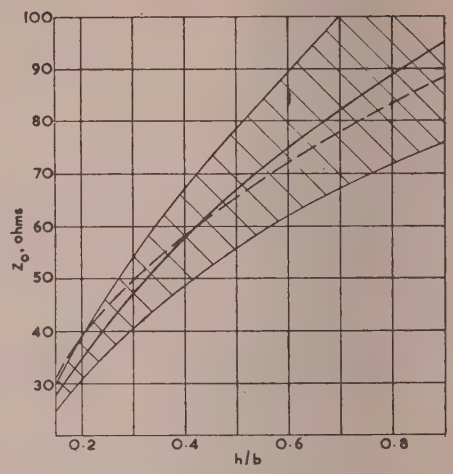


Fig. 21.—Comparison of measured and computed characteristic impedance for Teflon-fibre-glass Microstrip line.

—— Upper- and lower-bound solutions.
—— Pseudo-TEM solution, $a/b = 3$ to $a/b = \infty$.
---- Measured results at 7300 Mc/s.

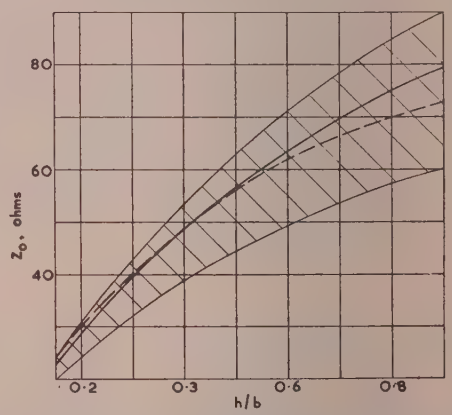


Fig. 22.—Comparison of measured and computed characteristic impedance for silicone-fibre-glass Microstrip line.

—— Upper- and lower-bound solutions.
—— Pseudo-TEM solution, $a/b = 3$ to $a/b = \infty$.
---- Measured results at 4770 Mc/s.

from eqns. (29) and (30) that the characteristic impedance is given by

$$Z_0 = 1/Cv \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

The disparity between measured and computed results is of the same general character for both λ_g/λ_0 and for Z_0. Furthermore, the cross-over or convergence point occurs at similar though not identical values of h/b. A more detailed inspection of the results suggests that the disparity can be accounted for almost entirely by an error in the estimate of capacitance. This, in turn, substantiates the belief that, for the dimensions commonly used and up to frequencies of at least 8000 Mc/s, the dominant mode may be regarded as a slightly perturbed TEM wave having phase velocity and characteristic impedance which do not differ significantly from those calculated by simple TEM theory. (For a further discussion of this point, see Section 13.4.)

For $h/b \geqslant 1{\cdot}0$, Arditi has concluded, on the basis of certain experimental evidence, that the wavelength tends to that value for a TM surface wave. For the two examples considered, and for the frequencies at which the specimens were tested, only the lowest-mode surface wave can propagate. The value of λ_g/λ

for this mode is approximately 0·9 in both cases.[26] The three sets of curves indicate that within the given range of h/b there is no apparent tendency for the dominant mode to degenerate into a surface wave. The question whether a surface wave can be simultaneously excited is, of course, an entirely different matter. Experimental evidence indicates that this can happen, especially if there exist any serious discontinuities in the line structure. Otherwise the magnitude of such spurious modes is usually sufficiently small to be neglected.

The same approach involving an upper-bound, a lower-bound and an intermediate pseudo-TEM solution may be applied to the calculation of attenuation. The upper and lower bounds are given directly by Figs. 16 and 17, but the pseudo-TEM attenuation can only be calculated by reference back to the basic current-distribution data together with a recalculation of the appropriate form factors. The intermediate pseudo-TEM approach also requires that the dielectric losses be estimated in such a manner as to allow for partial transmission of energy in the free-space region.

The complexity of such a calculation is not justified in the circumstances. First, the upper- and lower-bound solutions differ by an average of only 20% for the range of interest, and this is the order of divergence to be expected between theory and practice. Secondly, it is an experimental fact that the measured attenuation coefficient is a function of the distance from the launching device. The attenuation is a minimum in the immediate proximity of the launching device and rises asymptotically to a steady value at a distance of some 20 wavelengths. The significance of this phenomenon is not yet properly understood, but it is commonly encountered with open-wire and surface-wave transmission lines.

Once the steady state is reached, the measured attenuation is generally in reasonable agreement with theory, provided that the water absorption of the particular dielectric material used is reasonably low, otherwise serious discrepancies may be observed.[25] Thus, for the line illustrated in Fig. 2(*a*), with the dimensions shown, the attenuation at 4 000 Mc/s is approximately 1·0 dB/m, rising[5] to about 2·2 dB/m at 9 000 Mc/s. On substituting silicone fibre glass in place of p.t.f.e. fibre glass, the loss at 4 000 Mc/s increases to about 4 dB/m.

(8.2) Tri-plate Lines

The tri-plate line of Fig. 2(*a*) has an important advantage over the dielectric-supported strip-above-ground line of Fig. 2(*a*). The dielectric sheet is placed in the plane of symmetry where the electric gradient is, in theory at least, zero. Hence, provided that the sheet is sufficiently thin, its effect on the properties of the line can be neglected. In practice, a small reduction in impedance and a slight increase in attenuation are observed.[25] However, for low-loss dielectric materials such as p.t.f.e. fibre glass and pure styrene, and for the dimensions used in practice, the effect is small, and it is usually satisfactory to assume free-space conditions.

(9) MISCELLANEOUS PROPERTIES

(9.1) Coupling between Adjacent Strip Conductors

Measurements have been made with the aid of the electrolytic tank of the open-circuit capacitive coupling between two strip conductors possessing identical values of b and h above a common ground plane. A typical case is presented in Fig. 23, from which it will be observed that the coupling falls rapidly with spacing. Qualitatively this is in good agreement with experiments at microwave frequencies. These indicate that negligible interference occurs between circuits situated above a common ground plane, provided that they are spaced a few centimetres apart,

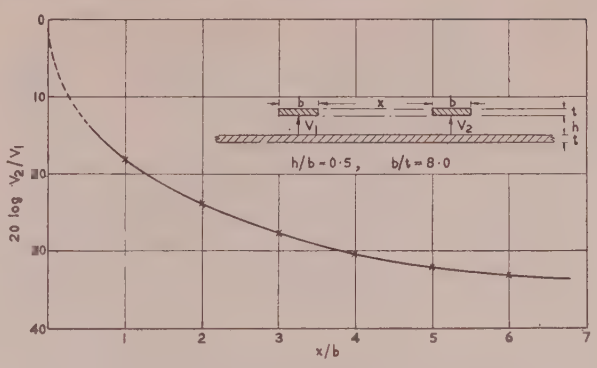


Fig. 23.—Open-circuit capacitive coupling between adjacent lines. V_1 is the excitation voltage, and V_2 is the induced voltage.

and that there are no severe structural discontinuities liable to cause radiation.

Some measurements were also made of coupling between adjacent strips in a tri-plate line, but they are not reproduced here because the conditions of measurement were not quite comparable. Nevertheless they support theoretical reasoning which suggests that, for identical values of b and h, the coupling is much less with the tri-plate line than with the strip above ground. At distances of $3b$ and over, the reduction in coupling is likely to exceed 30 dB.

In practice, a very low coupling factor is not necessarily an advantage. Even with the strip-above-ground system it is somewhat difficult to design a satisfactory 3 dB parallel-line directional coupler unless the coupling between the adjacent strip conductors is artificially enhanced in some way.

(9.2) Field at a Short Distance from the Transmission Line

Measurements have been made on the electrolytic tank of the electric gradient at a short distance (e.g. 3 times the strip width) from the centre of the strip transmission system for various configurations. On the average, the tri-plate line provides more efficient screening than the strip above ground. However, the polar plot of the tri-plate line shows a marked side-fire effect, the shape of the plot resembling the major lobes of a broadside antenna array. It seems probable that this effect is, in part, responsible for the severe outward radiation from the slot at right angles to the direction of propagation, which has been found to occur if the symmetry of the centre conductor is imperfect in the longitudinal direction. This tendency to radiate has been observed in a number of laboratories, and has proved very troublesome; it is most marked when $a \simeq n\lambda/2$. The simplest cure is to provide metallic posts connecting one ground plane to the other at intervals of less than $\lambda/4$ on either side of the strip conductor—a procedure which complicates manufacture and hinders the design of certain components, such as variable attenuators.

One further point worthy of mention concerns the relative field strength above and below a strip-above-ground system. Although for $a = 3b$ the impedance closely approximates to that for an infinite ground plane, the screening effect of a finite ground plane of this width is relatively poor. For example, for $h = 0{\cdot}27b$, $a = 3b$, $t = b/32$, the field at a distance $3{\cdot}5b$ below the centre of the ground plane is only 14 dB less than the field at the same distance above the centre of the strip conductor.

(10) CONCLUSIONS

The impedance and attenuation of air-spaced strip line have been derived from measurements on an electrolytic tank for a

range of parameters of practical interest. In so doing, it has been shown that certain earlier formulations for the impedance of a strip-above-ground system are seriously in error for values of Z_0 greater than 50 ohms, thus confirming the recent theoretical analysis by Black and Higgins. An approximate method has been evolved for calculating the pseudo-TEM characteristic impedance of a dielectric-slab-supported strip-above-ground system. This method gives results which are in reasonable agreement with calculations of characteristic impedance based on measurements of low-frequency capacitance and the velocity of propagation at microwave frequencies. This suggests that, for practical purposes, the dominant mode in the composite line is virtually TEM. However, this does not prejudice the question whether the line can at the same time support other modes.

Results have also been presented for the tri-plate line, thus facilitating a comparison with the strip-above-ground system. The tri-plate line appears to have advantages in respect of efficiency of screening, minimization of coupling between adjacent strip conductors (not always an advantage, however), minimization of dielectric losses, and ease of computation (since free-space conditions can in general be assumed). These advantages, however, are offset by certain practical limitations. Strictly speaking, these limitations do not lie within the scope of the paper, but in view of their importance they deserve a brief mention. First, unless strict symmetry of the strip conductor is maintained, particularly in the longitudinal direction, severe sideways radiation may occur. This may be overcome by providing metallic posts at frequent intervals which connect one ground plane to the other. This is at the expense, however, of considerable complications to the construction of the line. Secondly, the inaccessibility of the strip conductor is, by comparison, a severe disadvantage, particularly at the experimental stage.

Generally speaking, it is considered that, in those applications where printed strip line is an acceptable substitute for conventional hollow waveguide, the dielectric-slab-supported strip-above-ground system is likely to provide the most satisfactory arrangement. This is particularly true when considerable importance is attached to low weight, bulk and cost, together with ease of fabrication. Application of strip transmission lines to the design of microwave equipment will be described in a subsequent paper.[28]

(11) ACKNOWLEDGMENTS

Acknowledgment is made to Standard Telecommunication Laboratories for permission to publish the paper. The author wishes to thank Mr. T. R. Scott for encouraging the investigation described, and Mr. L. Lewin for a number of stimulating discussions.

(12) BIBLIOGRAPHY

(1) Grieg, D. D., and Engelmann, H. F.: "Microstrip—A New Transmission Technique for the Kilomegacycle Range," *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 1644.

(2) Assadourian, F.: and Rimai, E.: "Simplified Theory of Microstrip Transmission Systems," *ibid.*, p. 1651.

(3) Kostriza, J. A.: "Microstrip Components," *ibid.*, p. 1658.

(4) Arditi, M.: "Experimental Determination of the Properties of Microstrip Components" (Record of the I.R.E. National Convention, 1953, Part 10, Microwaves), p. 27.

(5) Arditi, M.: "Characteristics and Application of Microstrip for Microwave Wiring," *Transactions of the Institute of Radio Engineers*, March, 1955, **MTT-3**, p. 31.

(6) Sciegienny, J.: "Strip above Ground Plane Transmission System," *Quarterly Progress Report of the Research Laboratory of Electronics, M.I.T.*, Part X, January, 1953, p. 53.

(7) Sciegienny, J., and Schetzen, M.: "Strip Transmission Systems," and "Theoretical Analysis of a Strip Transmission System," *ibid.*, Part XI, April, 1953, p. 83.

(8) Schetzen, M.: "Printed Microwave Systems" (Technical Report 289, Research Laboratory of Electronics, M.I.T., September, 1954).

(9) Barrett, R. M.: "Etched Sheets serve as Microwave Components," *Electronics*, June, 1952, **25**, p. 114.

(10) Barrett, R. M.: "Microwave Printed Circuit—A Historical Survey," *Transactions of the Institute of Radio Engineers*, **MTT-3**, March, 1955, p. 1.

(11) Fubini, E., Fromm, W., and Keen, H.: "New Techniques for High-Q Strip-Microwave Components," *Convention Record of the I.R.E.*, 1954, Part 8, pp. 91–97.

(12) Fubini, E., Fromm, W., and Keen, H.: "Microwave Applications of High-Q Strip Components," *ibid.*, pp. 98–103.

(13) "Manufacture of Microstrip," *Electrical Communication*, December, 1952, **29**, p. 251.

(14) Packard, K. S.: "Machine Methods make Strip Transmission Line," *Electronics*, September, 1954, **27**, p. 148.

(15) Deschamps, G.: "Theoretical Aspects of Microstrip Waveguides," *Institute of Radio Engineers*, April, 1954, **PGMTT-2**, p. 100 (also **CS-2** No. 2, July, 1954, pp. 98–103).

(16) Montgomery, C. G., Dicke, R. H., and Purcell, E. M.: "Principles of Microwave Circuits," M.I.T. Radiation Series Vol. 8 (McGraw-Hill, New York, 1948), p. 22.

(17) Schelkunoff, S. A.: "Electromagnetic Waves" (Van Nostrand, New York, 1943), p. 179.

(18) Boothroyd, A. R., Cherry, E. C., and Makar, R.: "An Electrolytic Tank for the Measurement of Steady-State Response, Transient Response and Allied Properties of Networks," *Proceedings I.E.E.*, Paper No. 835 M, May, 1949 (**96**, Part I, p. 163).

(19) Boothroyd, A. R.: "Design of Electric Wave Filters with the Aid of an Electrolytic Tank," *ibid.*, Monograph No. 8, September, 1951 (**98**, Part IV, p. 65).

(20) Black, K. G., and Higgins, T. J.: "Rigorous Determination of the Parameters of Microstrip Transmission Lines," *Transactions of the Institute of Radio Engineers*, March, 1955, **MTT-3**, p. 93.

(21) Cohn, S. B.: "Characteristic Impedance of the Shielded Strip Transmission Line," *ibid.*, July, 1954, **MTT-2**, pp. 52–55.

(22) Cohn, S. B.: "Problems in Strip Transmission Lines," *ibid.*, March, 1955, **MTT-3**, p. 119.

(23) Dukes, J. M. C.: "The Application of Printed-Circuit Techniques at Microwave Frequencies," Paper in preparation.

(24) Zublin, K. E.: "Strip Type Components for 2000 Mc Receiver Head End," *Transactions of the Institute of Radio Engineers*, March, 1955, **MTT-3**, p. 65.

(25) Ringenbach, M. E., and Warren Cooper, H.: "Measurement of Attenuation and Phase Velocity of Various Laminate Materials at L-Band," *ibid.*, p. 87.

(26) Attwood, S. S.: "Surface Wave Propagation over a Coated Plane Conductor," *Journal of Applied Physics*, 1951, **22**, p. 504.

(27) Barlow, H. E. M., and Cullen, A. L.: "Surface Waves," *Proceedings I.E.E.*, Paper No. 1482 R, April, 1953 (**100**, Part III, p. 329).

[28] DAHLMAN, B. A.: "A Double-Ground-Plane Strip-Line System for Microwaves," *ibid.*, Paper No. 1852 R, July, 1955 (**102** B, p. 488).

(13) APPENDIX

(13.1) Basic Mathematical Relationships

(13.1.1) Low-Frequency Capacitance.

Let Fig. 24 represent the cross-section of an arbitrary two-conductor transmission line. It will be supposed that the two conductors are held in correct space relationship throughout

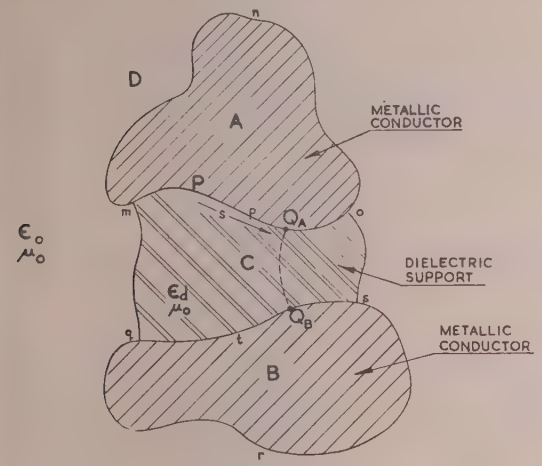


Fig. 24.—Arbitrary two-conductor transmission line with continuous dielectric support.
A. Metallic conductor.
B. Metallic conductor.
C. Dielectric support.

their length by a dielectric rod C. Let the shape of this rod be such that its edges *mq* and *os* lie along flux lines. Then the total charge carried on the surface *mpo* will equal that carried by the surface *qts*. Let these charges be designated, respectively, by q_p, q_t, q_n and q_r, instantaneous values per unit length being assumed. Let the instantaneous potential between the two conductors be v. The capacitance of the line is accordingly

$$C = \frac{q_p + q_n}{v} = \frac{q_t + q_r}{v} = C_{pt} + C_{nr} \quad . \quad . \quad (36)$$

i.e. the line behaves as though it consisted of two separate transmission lines in parallel, the one bounded by the region *mpostq*, and the other region *mqrson*. The fact that there is no interaction between the two lines is due, of course, to the assumption that the dielectric interfaces *mq* and *os* lie along flux lines.

Let s be the distances of a point Q_A on the surface of conductor A from an arbitrary reference point P. Let Q_B be the point on conductor B linked with Q_A by the same flux line. Let d be a distance on this flux line. Let $E(s)$ be the electric gradient at the surface of conductor A, assumed everywhere normal to the surface, and let $E(d)$ be the electric gradient along the flux line Q_AQ_B. At the point Q_A, $E(s)$ evidently equals $E(d)$.

Then the total charge on A is

$$q_p + q_n = \epsilon_0 \int_{mno} E(s)ds + \epsilon_d \int_{mpo} E(s)ds \quad . \quad . \quad . \quad (37)$$

The inter-conductor potential is

$$V = \int_{Q_AQ_B} E(d)\,dd \quad . \quad . \quad . \quad . \quad . \quad . \quad (38)$$

Hence

$$C = \epsilon_0 \left[\frac{\int_{mno} E(s)ds + \frac{\epsilon_d}{\epsilon_0} \int_{mpo} E(s)ds}{\int_{Q_AQ_B} E(d)\,dd} \right] \quad . \quad . \quad (39)$$

$$= \epsilon_0 M_d, \text{ say} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (40)$$

The factor M_d is evidently dimensionless, i.e. it is independent of actual physical size, and depends only on the geometry of the system. In order to obtain M_d it is necessary to know the functions $E(s)$ and $E(d)$. In principle, this can always be achieved by application of one of the several classical methods,[16, 17] but in practice, geometries of the kind considered in the paper cause very considerable difficulties.

The essential requirement is that the field satisfies everywhere the Laplacian conditions of eqn. (8). This must also be true of the two-dimensional analogue on the electrolytic tank, where A and B are conducting electrodes and the conducting sheet is divided into regions of resistivity ρ_0 and ρ_d. The solution is likewise given by

$$\frac{1}{R_d} = \frac{1}{\rho_0} \left[\frac{\int_{mno} E(s)ds + \frac{\rho_0}{\rho_d} \int_{mpo} E(s)ds}{\int_{Q_AQ_B} E(d)\,dd} \right] \quad . \quad . \quad (41)$$

where R_d is the measured resistance between the electrodes A and B.

Equating the two solutions for the given condition $\rho_0/\rho_d = \epsilon_d/\epsilon_0$, we have

$$C = \epsilon_0 M_d = \epsilon_0 \rho_0 / R_d \quad . \quad . \quad . \quad . \quad . \quad (42)$$

The calculation of capacitance is thus reduced to a simple measurement of resistance.

(13.1.2) Low-Frequency Inductance.

It is assumed that the permeability of the dielectric material is equal to that of free space. The presence of the dielectric region C may therefore be ignored. Applying the same methods as before, the following result is obtained for the loop inductance of a line of unit length:

$$L = \mu_0 / M_0 = \mu_0 R_0 / \rho_0 \quad . \quad . \quad . \quad . \quad . \quad (43)$$

where $M_0 = M_d$ for $\epsilon_d = \epsilon_0$
$R_0 = R_d$ for $\epsilon_d = \epsilon_0$

(13.1.3) Low-Frequency Impedance and Velocity of Propagation.

For frequencies approaching zero it is assumed that there is no interaction between L and C. Hence the impedance and velocity of propagation will be given by

$$\bar{Z}_0 = (L/C)^{1/2} = [\mu_0/(\epsilon_0 M_0 M_d)]^{1/2} = (\mu_0 R_0 R_d/\epsilon_0)^{1/2}/P_0 \quad (44)$$

$$\bar{v} = (LC)^{-1/2} = [M_0/M_d \mu_0 \epsilon_0]^{1/2} = c(R_d/R_0)^{1/2} \quad . \quad (45)$$

For the condition $R_d = R_0$, i.e. $\epsilon_d = \epsilon_0$, eqn. (44) reduces to eqn. (4).

For $\epsilon_d \neq \epsilon_0$ the method described in Section 8.1 may be used to define the impedance in terms of upper- and lower-bound solutions. The lower bound (corresponding to lower impedance) is obtained by setting $\epsilon = \epsilon_d$ throughout the entire cross-section. In this case eqn. (39) becomes

$$C_l = \epsilon_d \left[\frac{\int_{mno} E(s)ds + \int_{mpo} E(s)ds}{\int_{Q_AQ_B} E(d)\,dd} \right] \quad . \quad . \quad . \quad (46)$$

Comparing eqns. (39) and (46) it will be apparent that this corresponds to the situation $\epsilon_d \to \epsilon_0$.

The upper-bound solution is given by assuming that $\epsilon_d \gg \epsilon_0$, in which case

$$C_u = \epsilon_d \left[\frac{\int_{mpo} E(s)ds}{\int_{Q_A Q_B} E(d)\,dd} \right] \quad . \quad . \quad . \quad . \quad (47)$$

Let $$C_u/C_l = \nu^2, \text{ say} \quad . \quad . \quad . \quad . \quad . \quad (48)$$

If the corresponding upper- and lower-bound solutions are Z_u and Z_l, then, from eqn. (44),

$$\nu = Z_l/Z_u$$

which is independent of the value of ϵ_d and corresponds to eqn. (34).

Substituting eqns. (46)–(48) in eqn. (39) it will be found that

$$\frac{C}{C_l} = \frac{\epsilon_0}{\epsilon_d}\left[1 + \nu^2\left(\frac{\epsilon_d}{\epsilon_0} - 1\right)\right] \quad . \quad . \quad . \quad (50)$$

Let $\left[1 + \nu^2\left(\frac{\epsilon_d}{\epsilon_0} - 1\right)\right]^{-1/2} = \xi$ as in eqn. (33)

Then by substitution in eqn. (44), the impedance of the composite line will be given by

$$\bar{Z}_0 = \left(\frac{\epsilon_d}{\epsilon_0}\right)^{1/2} \xi Z_l \quad . \quad . \quad . \quad . \quad (51a)$$

$$= \xi Z_0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (51b)$$

where Z_0 is the solution for the line of Fig. 24, as given by eqn. (44) for the specific condition $\epsilon_d = \epsilon_0$.

It should be stressed that this method of specifying the characteristic impedance is entirely rigorous under the following conditions:

(*a*) The dielectric interfaces lie along flux lines.
(*b*) The wavelength is very large compared with the physical dimensions of the line.

When the upper and lower bounds are considered to be the two lines shown in Figs. 18(*a*) and 18(*b*), the intermediate solution [for Fig. 19(*b*)] ceases to be rigorous, even though the upper and lower bounds are exactly determined. This is because the dielectric interfaces of Fig. 19(*b*) do not lie along the stream lines.

(13.1.4) **Attenuation.**

Let δI be the current carried at Q_A in Fig. 24 by an elementary strip of width δs and of unit length in the direction of propagation.

Then $$\delta I = H(s)\delta s \quad . \quad . \quad . \quad . \quad . \quad . \quad (52)$$

where $H(s)$ is assumed to be tangential to the surface at Q_A. $E(s)$, which is assumed to be normal to the surface, and $H(s)$, are related by the wave impedance

$$E(s) = Z_\eta H(s) \quad . \quad . \quad . \quad . \quad . \quad (53)$$

The r.f. skin resistance of the elementary strip is given by

$$\delta R = R_s \delta s \quad . \quad . \quad . \quad . \quad . \quad . \quad (54)$$

where $$R_s = (\pi f \mu/\sigma)^{1/2} \quad . \quad . \quad . \quad . \quad . \quad (55)$$

Hence the loss in the elementary strip, assuming r.m.s. values, is given by

$$\delta W = \delta R(\delta I)^2 \quad . \quad . \quad . \quad . \quad . \quad (56a)$$

$$= \frac{R_s}{Z_\eta^2} E^2(s)\delta s \quad . \quad . \quad . \quad . \quad . \quad (56b)$$

Likewise the total loss in both conductors A and B, per unit length, is

$$\frac{R_s}{Z_{\eta d}^2}\left[\int_{mpo} E^2(s)ds + \int_{qts} E^2(s)ds\right] + \frac{R_s}{Z_{\eta 0}^2}\left[\int_{mno} E^2(s)ds + \int_{qrs} E^2(s)ds\right] \quad . \quad . \quad . \quad . \quad (57)$$

where $Z_{\eta d}$ and $Z_{\eta 0}$ are, respectively, the wave impedances in the dielectric and free-space regions.

If V is the potential between the lines, the power transmitted will, by the same reasoning, be given by

$$\frac{V}{Z_{\eta d}}\int_{mpo} E(s)ds + \frac{V}{Z_{\eta 0}}\int_{mno} E(s)ds \quad . \quad . \quad . \quad (58)$$

The attenuation of the line (metallic losses only) is therefore given by

$$\alpha_m = \frac{R_s}{2V} \frac{Z_{\eta d}^{-2}\left[\int_{mpo} E^2(s)ds + \int_{qts} E^2(s)ds\right] + Z_{\eta 0}^{-2}\left[\int_{mno} E^2(s)ds + \int_{qrs} E^2(s)ds\right]}{Z_{\eta d}^{-1}\int_{mpo} E(s)ds + Z_{\eta 0}^{-1}\int_{mno} E(s)ds} \quad . \quad (59)$$

For the air-spaced line $\epsilon_d = \epsilon_0$, and this equation reduces to the simple form given in eqns. (5) and (7). (The factor of one-half occurs because the attenuation constant α_m is the exponent of *voltage* or *current* decay.)

The electrolytic tank, in providing the function $E(s)$, allows calculation of the above integral. However, as has been stressed in Section 3, it is more convenient to measure the distribution of currents flowing out into the electrolyte rather than the surface potential gradient.

The method of upper and lower bounds may also be applied to the calculation of attenuation, but the numerical computations are considerably more tedious.

(13.1.4) **Validity of the TEM Approach for Composite Lines.**

It is questionable to what extent the TEM approach is valid when there exists a dielectric interface between two media possessing different permittivities. This proposition follows directly from the application of the wave equation

$$\nabla^2 E + \mu\epsilon\ddot{E} = 0 \quad . \quad . \quad . \quad . \quad . \quad (60)$$

which must be satisfied at all points throughout both media.

For sinusoidal waves of angular frequency ω this equation may be rewritten

$$\frac{\partial^2 E_x}{\partial x^2} + \frac{\partial^2 E_x}{\partial y^2} + (\omega^2\mu\epsilon - \beta^2)E_x = 0 \quad . \quad . \quad (61)$$

where it is assumed that the wave is travelling exclusively in the z-direction with a phase constant β. Two other identical equations must be satisfied for E_y and E_z.

For TEM propagation $\beta = \omega\sqrt{(\mu\epsilon)}$, and hence the solution satisfies the Laplacian requirement for electric distributions in the xy-plane, namely

$$\frac{\partial^2 E_x}{\partial x^2} + \frac{\partial^2 E_x}{\partial y^2} = 0, \text{ and } \frac{\partial^2 E_y}{\partial x^2} + \frac{\partial^2 E_y}{\partial z^2} = 0 \quad . \quad . \quad (62)$$

If the permittivity ϵ has one value in one medium and another value in the other medium, β will have two values, one for each medium. However, this is manifestly absurd since one wave can only possess one phase velocity. From this it follows that if the wave is to propagate uniformly along the line without radiation there must exist—as a minimum—a component of the electric field in the direction of propagation.

In view of the similarity in boundary conditions between the dielectric-supported strip-above-ground line and an image-type dielectric guide (i.e. one which is divided on its plane of symmetry by a conducting sheet), it is probable that the dominant mode is of the EH-type. This appears to be confirmed by the work of Sciegienny and Schetzen.[7,8]

On the other hand, it is apparent that most of the power is likely to be concentrated in the region between the strip conductor and the ground plane, and in this region the field should be substantially TEM. This suggests that the solutions based on the TEM approach are likely to give a reasonable approximation provided that the physical dimensions are small; this is in accordance with experimental observation.

The character of the fringe fields is best understood by referring once again to the phenomena of surface-wave transmission. These suggest that, within the dielectric slab, E_x and E_y will vary harmonically, and that in the external free-space region they will decay exponentially, coupled with a progressive phase advance, with increasing distance. However, it is almost certain that, owing to the complex boundary conditions, this harmonic variation would be such as to involve an infinite set of eigen values *for any single mode*. This situation is akin to space harmonics in periodic structures, where each individual space harmonic does not by itself satisfy the boundary conditions.

Barlow and Cullen[27] have provided a convenient physical structure of surface-wave transmission which is of relevance in a qualitative manner to the present problem. These authors show that the Zenneck wave may be represented by an incoming plane wave incident at the Brewster angle. A similar situation must exist in the present case, except that the E vectors will be bent in closed loops passing between the strip conductor and ground plane. Above the strip conductor they will, however, have the characteristic forward tilt of surface-wave transmission. The physical picture of an incoming wave is additionally useful in connection with a discussion of launching devices. In order to stimulate such a wave, the aperture of the launching device would have to be of infinite size. As in practice it is not, it is to be expected that the proper field conditions will only be established at some distance along the line from the launching device. The variation in measured attenuation coefficient, mentioned briefly in Section 8.1, is not therefore an altogether surprising phenomenon.

(13.2) Sources of Error in Electrolytic-Tank Measurements

Provided that the electrodes were cleaned regularly, that the electrode surface current density was not allowed to exceed 0·3 mA/mm^2, and that the duration of a single set of measurements was not excessive, relatively little trouble was experienced due to corrosion. Surface currents approaching, but not exceeding, the above maximum were conveniently obtained with a weak solution of about 2 g of copper sulphate per litre of distilled water, with inter-electrode voltages ranging from 6 to 24 volts. The effect of corrosion, which is to set up a variable electrolytic capacitance between the electrode and the electrolyte, can be minimized by employing a sufficiently high supply frequency.[18] In the present instance, however, the area of contact was so large that no difficulty was experienced in operating at a frequency of 50 c/s.

There are, in effect, two possible sources of error. The first is due to the finite gap between two segments, which does not collect or deliver current to the electrolyte. Some tests were therefore carried out to see whether the segments collected the current that otherwise would have been collected by the gap. Except when two electrodes were very close to each other, this was found to be the case. Hence, in general, it may be assumed that the segmented electrode has a behaviour identical to that of a solid metallic electrode, and that the collection area of the segment is equal to one segment plus two half gaps. The second error due to segmentation arises as a result of the sampling error in regions such as the extremity of the electrode, where the current density is changing rapidly. This subject is discussed in Section 7.2, and a curve (see Fig. 14) was obtained which enables an empirical approximate correction to be made to the computed form factor.

Owing to the miniscus effect, the electrolyte climbs up the surface of the electrode to a height nearly double the depth of the electrolyte layer. At the extremity of the electrode the sharp corners cause a considerable variation in this height. However, tests showed that a very considerable variation in wetted area could be tolerated before there was any appreciable change in current reading. In fact, it was possible to lift a complete electrode almost out of the solution before any significant change was observed. Once again, however, it is to be expected that an error might be introduced for very close electrode spacings —closer than were used in the present instance.

The average change in conductivity during a working day was about 5%, and for this reason it was found necessary to check the conductivity every few hours. Furthermore, the evaporation was sufficient to necessitate daily topping-up of the tank. (The additional fluid must be well mixed with that already in the tank.)

The effect of the ammeter impedance (neglecting reactance) is to lower the voltage on the segment under test relative to the adjacent segments, which are all connected to the supply source. In Fig. 25, R_1 is the total resistance to all the segments at the same potential, and R_2 is the total resistance to segments of opposite polarity. If the voltage drop in the ammeter is very small compared with V_1, we may say that the quantity we require to know is i_2, which may be expressed in terms of the quantity we actually measure, namely i, by means of the relationship

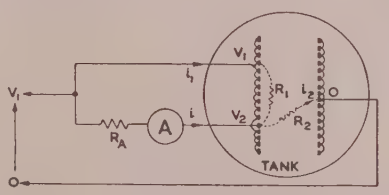


Fig. 25.—The effect of ammeter internal resistance.

$$i_2 = i(1 + R_A/R_1) \quad . \quad . \quad . \quad . \quad . \quad (63)$$

Hence for the error to be small the ammeter resistance must be small compared with the total inter-segment resistance. In practice this is very difficult to achieve, and values of about 100 ohms and 500 ohms, respectively, are to be expected. On the other hand, if the ratio is reasonably constant for every segment, every reading will be altered proportionately, and the form factor will not be affected. Despite the variation in path length at the corners of the electrode, the inter-segment resistance is surprisingly constant, and in the worst case the variation in $(1 + R_A/R_1)$ was found to be less than 6%. The net effect on the calculated attenuation is smaller than this, however—a fact which was checked by introducing a special circuit which raised the potential of the segment under test to that of its neighbours. The net error (due to this cause) for the types of geometry considered would appear not to exceed a total of 2%.

621.385.5.032.75

Paper No. 2038 M
May 1956

ELECTRON BOMBARDMENT OF THE GLASS ENVELOPE OF A RECEIVING VALVE

By G. H. METSON, M.C., Ph.D., M.Sc., B.Sc.(Eng.), Associate Member, and D. J. SARGENT.

(*The paper was first received 26th September, 1955, and in revised form 10th January, 1956.*)

SUMMARY

The probable causes of the development of stable bombarding potentials on the glass envelope of an oxide-cathode pentode are analysed and subjected to experiment. A photo-electric effect is examined in some detail and shown to give rise to dangerous potentials under certain conditions which may be of interest to electronic switching engineers.

(1) INTRODUCTION

The general phenomenon connected with the electron bombardment of glassware in receiving valves will be familiar to most circuit engineers. Its occurrence leads to degraded performance in the h.f. stage of amplifiers, to serious distortion in the output stage and probably to shortened valve life. The course taken during the build-up of the bombarding condition may be described in the following manner. During the operation of the valve certain parts of the insulating system—mica or glass—may acquire a positive potential. This potential attracts electrons from the electrode system, and the resulting insulator bombardment leads to one or other of two distinct states: if the insulator potential is such that its secondary-emission coefficient, δ, is less than unity the insulator will collect electrons and its potential will fall to some value approaching cathode potential; however, if the potential is sufficiently high for δ to exceed unity the insulator will lose electrons and its potential will approach the anode voltage of the valve. This condition, described in the paper as a 'lock-up', is quite stable and can be broken down only by disconnecting the anode voltage for a time sufficient to allow the charge on the insulator to leak away to earth, so that its potential is lowered to a value where δ is less than unity. These matters have been described by Benjamin, Cosgrove and Warren,[1] McKay,[2] Koller,[3] Deketh,[4] and others.

A typical characteristic of δ with respect to electron bombarding energy for an insulator is shown in Fig. 1. There are two

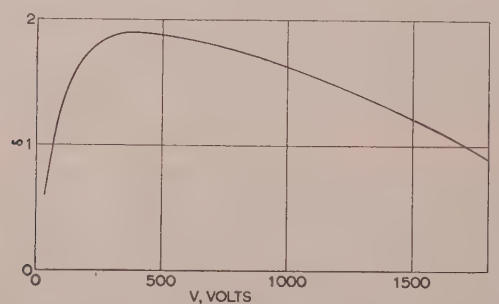


Fig. 1.—Typical (secondary emission coefficient)/(primary electron energy) characteristic for an insulator.

energies at which $\delta = 1$ and these are usually known as 'sticking' potentials. McKay quotes the following values for the lower and upper sticking potentials for various glasses and for barium oxide (quoted because the interior surfaces of the glass envelopes of most receiving valves are contaminated with barium oxide).

Material	Lower sticking potential	Upper sticking potential
	volts	volts
Pyrex glass ..	40	2300
Cover glass ..	60	1700
Quartz glass ..	30	2300
Barium oxide ..	60	3500

Most receiving-valve envelopes are made from soda and lead glasses, and the values for these are not quoted by McKay. The authors have, however, noted the occurrence of the lower sticking potential in conventional receiving valves over the range 40–100 volts.

(1.1) Object of the Experimental Work

Technical interest in the lock-up phenomenon centres on the manner in which the insulator acquires a positive potential such that $\delta > 1$, and it is usually assumed that the potential build-up is due to one or both of the following causes:

(*a*) Leakage of electrons to the anode over insulating surfaces, resulting in a rising positive potential on mica or glassware.
(*b*) Capacitive transfer of potential from the anode to mica and glass surfaces.

Recent work at the Post Office Research Station on the mechanism of build-up has found it to be complex and probably due in common valves to the additive effect of a number of causes. Anode leakage and capacitive transfer of voltage play an important part, but a further effect due to a photo-electric emission phenomenon may be of equal or even greater importance under particular conditions.

The paper describes the effects leading to the attainment of the lower sticking potential and examines the conditions under which a lock-up state may occur. The case of anode leakage has been omitted, as it is self-evident.

(1.2) The Experimental Valve

The valve type selected for experiment was a high-gain indirectly-heated oxide-cathode pentode with an open parallel-plate anode system. Direct anode-voltage leakage to the glass immediately surrounding the electrode system was avoided by the use of micas too small to touch the envelope at any point. Leakage of anode voltage from the pinch was checked by a guard-ring system. This complete elimination of anode-voltage leakage to the glass enabled other components of the glass potential to be separated and measured in isolation.

(2) FACTORS INFLUENCING THE POTENTIAL OF THE GLASS ENVELOPE

(2.1) Factors Tending to Increase the Potential

During operation of the valve the glass envelope tends to increase in positive potential, owing to the following five causes:

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Dr. Metson and Mr. Sargent are at the Post Office Research Station.

(*a*) Anode-voltage leakage, I_l. This is mainly due to surface leakage over mica and glass.

(*b*) Capacitive transfer of anode voltage, V. The glass adjacent to the anode may be regarded as the juncture of two capacitances, of which the outer elements are connected to anode and earth respectively. The glass therefore takes up an intermediate potential depending on the relative capacitances.

(*c*) Photo-electric emission, I_p. In oxide-cathode valves the inner wall of the glass envelope is contaminated with barium evaporated from the cathode. Normal illumination causes this surface to emit photo-electrons which pass spatially to the anode. Loss of this electron stream leads to a rising positive potential on the glass.

(*d*) Soft X-ray irradiation, $k_0 I_a$. The electron-bombarded anode emits soft X-rays, which irradiate the glass wall of the envelope and cause it to emit photo-electrons. The magnitude of this stream is $k_0 I_a$, where k_0 is the residual vacuum factor[5] of the valve and I_a is the anode current.

(*e*) Ionic flow, I_i. A certain number of positive ions of residual gas will flow to the glass envelope and be equivalent to the loss of an equal number of electrons. This flow is negligible in well-processed valves but may be of importance during the manufacturing processes.

(2.2) Factors Tending to Decrease the Potential

Two factors tend to lower the potential of the glass during valve operation, namely

(*a*) Leakage to earth or cathode, I_e. A stream of electrons passes over the surface and through the volume of the glass from the cathode to the section of envelope adjacent to anode. This stream partially replaces the electron losses mentioned in Section 2.1 and tends to reduce the potential.

(*b*) Low-voltage electron bombardment, $I_f(1-\delta)$. As soon as the inner wall of the envelope assumes a positive potential some free electrons will move out of the electrode system proper and bombard the glass. Provided that the energy of impact is such that $\delta < 1$ this bombardment will lower the potential.

The various components detailed in Sections 2.1 and 2.2 are shown in Fig. 2.

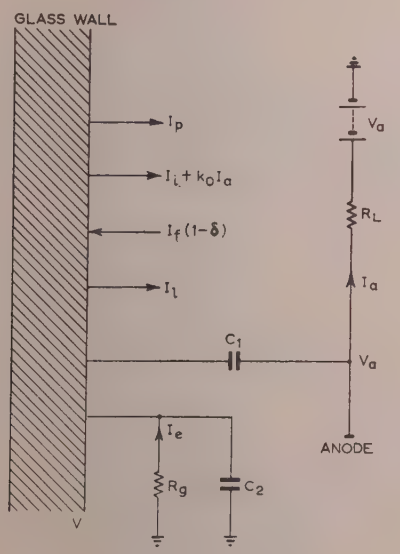


Fig. 2.—Circuit analysis.

(3) TENTATIVE CIRCUIT ANALYSIS

The derivation of the glass potential from the components shown in Fig. 2 is complex, but partial solutions might be approached in the following manner. Suppose the voltage V_a is suddenly applied to the anode through the anode load R_L. The capacitances C_1 and C_2 become charged almost instantaneously, since they are small (1–5 $\mu\mu$F) and R_L is usually a few thousand ohms. The peak voltage appearing on the glass, V, will approach

$$V = V_a \frac{C_1}{C_1 + C_2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

and then slowly discharge to earth potential through the glass leakage resistance, R_g. The rapid charge and slow discharge are due to the relative magnitudes of R_L and R_g, which may have a ratio of the order of $1 : 10^6$. The transfer voltage, V, in eqn. (1) may therefore be regarded as essentially transient, appearing whenever V_a changes. Owing to earth leakage the voltage will be dependent on the rate of change of V_a and will achieve its maximum value after an instantaneous switching operation.

The remaining components in Fig. 2 can be compounded on a steady-state basis:

$$\begin{aligned} V &= R_g I_e \\ &= R_g[I_p + I_i + k_0 I_a + I_l - I_f(1-\delta)] \quad . \quad . \quad (2) \end{aligned}$$

For the case in hand, eqn. (2) can be simplified, since the ionic flow, I_i, is negligible in well-processed valves and the anode leakage, I_l, has been eliminated in the experimental valve. Thus,

$$V = R_g[I_p + k_0 I_a - I_f(1-\delta)] \quad . \quad . \quad . \quad . \quad (3)$$

The voltage components on the glass due to the photo-electric and X-ray effects can now be separated and derived from eqns. (2) and (3). For example, if the valve is operated at an anode voltage V_a and back-biased on the control grid to a condition of zero anode current, the glass potential can assume one or other of the following values:

$$\left.\begin{aligned} V &= 0 \\ \text{or} \quad V &= R_g I_p \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The first condition occurs when the valve operates in darkness ($I_p = 0$) and the second when under illumination. It will be shown later that this photo-electric component of voltage may be very considerable and raise the glass potential well above the level where $\delta > 1$. Derivation of the potentials arising from X-ray irradiation will be deferred to Section 6.

This brief survey is hardly rigorous, but it forms a useful introduction to the experimental work discussed in subsequent Sections.

(4) CAPACITIVE TRANSFER OF ANODE VOLTAGE

(4.1) Magnitude of Quantities Involved

Fig. 2 shows the layout of capacitances involved in the transfer. The potential, V, of the glass for an anode potential of V_a is

$$V = V_a \frac{C_1}{C_1 + C_2}$$

Measurement of the capacitances of typical samples of the experimental valve gave C_1 as 0·6 $\mu\mu$F and C_2 as 3·0 $\mu\mu$F, whence $V = 0{\cdot}16 V_a$.

This relationship is shown in Fig. 3. Direct measurement of the V/V_a characteristic can be made on a quadrant electrometer, provided that the earth leakage, I_e, can be eliminated. Such a measurement was attempted on a special valve in which R_g was made infinitely high by an air-gap. The capacitance of the electrometer must, however, be reduced by a series capacitor to a value small compared with C_2, and this leads to loss of accuracy, owing to insensitivity of the electrometer. A single experiment conducted on these lines gave a characteristic similar to that shown in Fig. 3 but about 30% lower in magnitude.

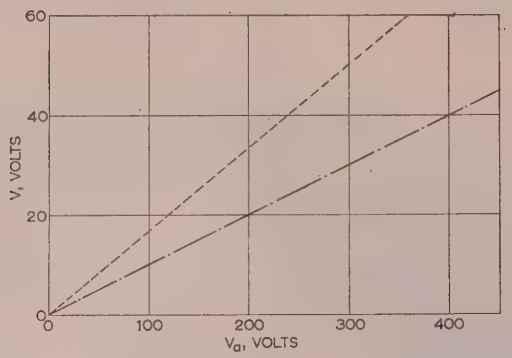


Fig. 3.—(Glass potential)/(anode potential) characteristics.
----- Calculated from $V = 0{\cdot}16V_a$.
—·—· Measured with electrometer ($C_e = 0{\cdot}3\ \mu\mu F$).

With a minimum of 40 volts for the lower sticking potential on soda-glass envelopes, Fig. 3 indicates the improbability of developing a lock-up condition in the experimental valve type below an anode potential of about 200 volts.

(4.2) Measurement of V_a for Sticking Potential due to Capacitive Transfer

It will be shown in Section 6 that $k_0 I_a$ plays no effective part in the development of sticking potentials in practical valves. The experimental valve was specifically designed to eliminate the anode leakage component, and it can therefore be used for direct investigation of the capacitively induced sticking potential, provided that the photo-electric component is eliminated by operating the valve in darkness. The measurement is carried out in the following manner. The valve is provided with a narrow Aquadag band around the exterior of the glass envelope adjacent to the anode system. This band is normally free, but it can be connected at will either to earth for discharging the glass system or to a quadrant electrometer for potential measurement. The anode and screen voltages are first set at some low value (100 volts) and the anode current is adjusted to 5 mA. The anode voltage is now switched off, the Aquadag band being first connected to earth for discharge and then insulated. Finally, the anode voltage is reconnected by a switching operation and the Aquadag band is then connected to the electrometer for a potential measurement. If this potential is within a volt or two of earth level, no lock-up condition has developed. The whole process is now repeated at a slightly higher V_a until the electrometer shows a sudden rise from earth potential to some value approaching V_a.

Such tests were carried out on a batch of 20 valves and showed that the sticking condition developed over an anode-voltage range of 250–400 volts. Several useful points emerged from these tests. It was found, for example, that the same valve always developed the sticking condition at approximately the same anode voltage, irrespective of the magnitude of I_a. Measurement of C_1 and C_2 showed a considerable uniformity over the series, and variation of the slope of the calculated V/V_a characteristic could not explain the wide variation of V_a at which the sticking condition set in. It was therefore concluded that the variation was due to the state of the bombarded glass surface, i.e. a variation of V for the condition $\delta = 1$.

The tests were repeated with a modification: the switching operation was eliminated and V_a was advanced slowly and smoothly from 100 to 400 volts at constant anode current. ! No sticking condition was observed in any case, and this result is in conformity with the observations in Section 3.

(5) THE PHOTO-ELECTRIC EFFECT

The inner surface of the glass envelope of an oxide-cathode valve is usually contaminated with evaporated barium metal and shows considerable photo-electric activity under external illumination. The emitted photo-electrons are drawn to the anode and the glass potential rises in the manner previously recorded ($V = R_g I_p$).

Direct measurement of the effect can be made on working valves if a quadrant electrometer is connected to a narrow Aquadag band painted on the outside of the glass envelope adjacent to the anode system. To effect a measurement the anode current must be biased back to zero and a characteristic of V/V_a taken. The reduction of I_a to zero is, of course, necessary to ensure elimination of I_f [the same result may be obtained by measuring V/V_a with the cathode in a non-emitting state ($V_h = 0$)]. Some typical results on a group of experimental valves are shown in Fig. 4. The illumination used for the

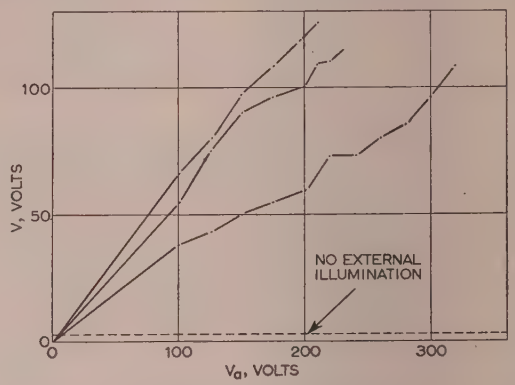


Fig. 4.—Relationship between glass potential due to photo-electric effect and anode potential.

measurements was that normally found at bench level in a well-lit laboratory. The broken line in Fig. 4, showing a constant 2–3 volts over the whole range of V_a, is typical of each of the valves when the characteristic is taken under conditions of complete external darkness. This result shows that the light from the cathode itself (970–1 000° K) is insufficient to produce a tangible photo-electric effect.

Comment may be useful at this stage on the influence of R_g on the glass potential. If R_g is zero the glass clearly remains at earth potential and no increase in V_a or intensity of illumination can affect it. However, if R_g is infinitely high, V_g will rise steadily with loss of photo-electrons, until it is approximately equal to V_a. No further rise is then possible, since the photo-electrons will find themselves in a retarding field and will therefore return to their point of origin. The slopes of the V/V_a characteristics in Fig. 4 may thus be regarded as functions of the individual values of R_g.

The anode-voltage rating of the valve type shown in Fig. 4 is 300 volts. If the anode voltage is applied before the heater voltage, so that the photo-electric glass potential can build up undisturbed by I_f, the glass in each case will be at a potential greater than 35 volts when free electrons are ultimately released from the electrode system. Under such conditions a lock-up condition is likely to develop to the glass, and practical tests showed that, in fact, each valve reacted in the manner predicted.

(6) SOFT X-RAY EFFECT

(6.1) Magnitude of the Effect

The X-ray effect has much in common with the photo-electric effect. Both involve the emission of photo-electrons from the

glass and the subsequent passage of this electron stream to the anode. The X-ray case, however, is more complex, since it can arise only in the presence of an anode current, which automatically implies the existence of a free-electron stream. This stream has an important bearing on the course taken by the effect and this will be considered later.

If the anode of the valve is bombarded by a current I_a under a potential V_a, X-rays from the anode will irradiate the glass and cause a certain photo-emission of electrons of magnitude $k_0 I_a$ to pass from the glass to the anode. The factor k_0 is the residual vacuum factor of the valve with respect to the glass envelope and is completely defined by the valve geometry, the anode voltage and the anode material.[5,6]

The magnitude of k_0 follows from eqn. (2):

$$I_e = k_0 I_a - I_f(1 - \delta) \quad . \quad . \quad . \quad . \quad . \quad (5)$$

when the experimental valve is used in darkness ($I_l = I_p = 0$). An experimental aid can be introduced at this point by employing the external Aquadag band around the envelope to collect the current I_e. Owing to the envelope geometry, the resistance between the inner wall of the envelope and the Aquadag band is very small compared with the earth leakage path R_g, and effectively the whole of I_e is directly measurable on a galvanometer connected between the Aquadag band and earth.

The passage of the current I_e through the glass produces a gradient which raises the inner surface to some value above earth potential. Free electrons are now attracted to the inner surface, and its potential falls to an intermediate value. If a battery of dry cells is included in the external galvanometer circuit, with the negative pole connected to the Aquadag band, the p.d. through the glass can be balanced out and the inner surface lowered to earth potential, whereupon I_f tends to zero. Further increase of the magnitude of the external battery forces the interior surface progressively more negative and eliminates I_f without affecting $k_0 I_a$. The current through the galvanometer is now $I_e = k_0 I_a$, which is dependent only on X-ray quanta and is independent of the voltage of the external battery or the resistance of the glass. A typical characteristic of $k_0 I_a$ is shown in Fig. 5 for the experimental valve, and the basic k_0/V_a is, of course, one-tenth of this, since the characteristic has been drawn for the particular case $I_a = 10$ mA. The characteristics for individual valves show little variation from the quoted example. The reason for the discontinuities in the characteristic has been suggested elsewhere.[6]

The photo-electric current for a typical valve under normal artificial laboratory illumination is also shown in Fig. 5. Depending primarily on light quanta, it is, of course, almost independent of anode voltage. During normal valve operation both components of current, I_p and $k_0 I_a$, will be drawn through the common resistance R_g, and the resulting voltage components will therefore be proportional to the current components. Above $V_a = 260$ volts it is apparent that the X-ray component becomes progressively greater than the optical component. Fig. 4 shows that the optical component is usually greater than 80 volts, and it follows that the X-ray component should result in even greater magnitudes for $V_a > 260$ volts.

(6.2) Course of the Effect

The course taken by the effect can be examined by observing the variation of anode voltage at which a sticking condition develops with change of anode current. Numerous experiments failed, however, to show any such change, and V_a for a sticking condition was always found to be invariant with anode current over a wide range of 1–20 mA. Reasons for this rather surprising result are put forward in the next Section.

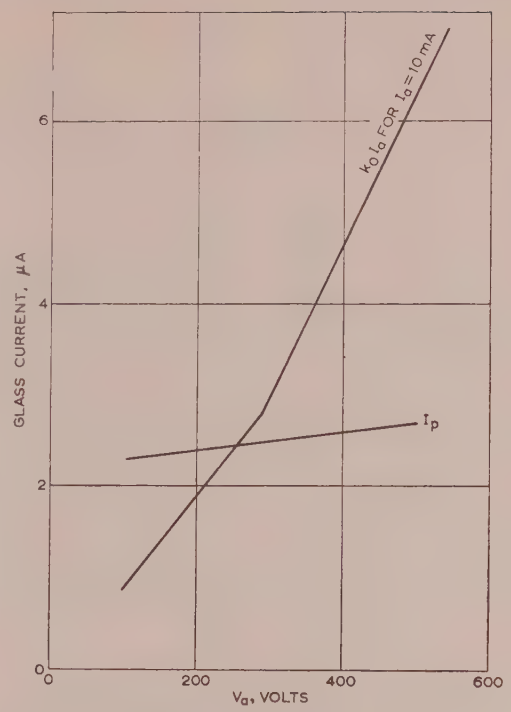


Fig. 5.—Relative magnitudes of photo-electric and X-ray effects for the experimental valve.

(7) DEVELOPMENT OF THE LOCK-UP CONDITION

Brief consideration will now be given to the manner in which the lock-up condition appears to build up. Certain transfer effects in the valve tend to increase the positive potential of the glass. These include a transient effect due to capacitive voltage transfer, together with certain steady-state effects such as anode leakage and radiation effects. Under favourable circumstances these effects, either separately or in combination, enable the glass potential to rise to a value where $\delta > 1$ and a lock-up condition occurs; in other circumstances the glass potential is clamped to something approaching earth potential by the arrival of low-energy electrons escaping from the electrode system. These low-energy electrons, I_f, are of basic importance in determining the course taken by the glass potential. The crucial point appears to be the time at which they arrive with respect to the rising glass potential. If arrival coincides with the condition $\delta < 1$ the glass returns to earth potential; if $\delta > 1$ the glass potential mounts swiftly to a stable level approaching that of the anode.

(7.1) Low-Energy Bombardment of Glass

The number of electrons leaving the electrode system proper and striking the glass is primarily a function of glass potential V. Fig. 6 shows a series of characteristics relating the potential of the inner glass surface, V, to the bombarding current, I_f for different anode currents in the experimental valves. For this particular measurement the inner surface of the envelope was coated with an Aquadag band connected to the exterior of the valve by a wire passing through the envelope. The potential of the inner surface was controlled by an external battery and the current was measured with an earth-connected microammeter.

The characteristics show that with negative values of V the magnitude of I_f is negligibly small; it is, in fact, equal to the

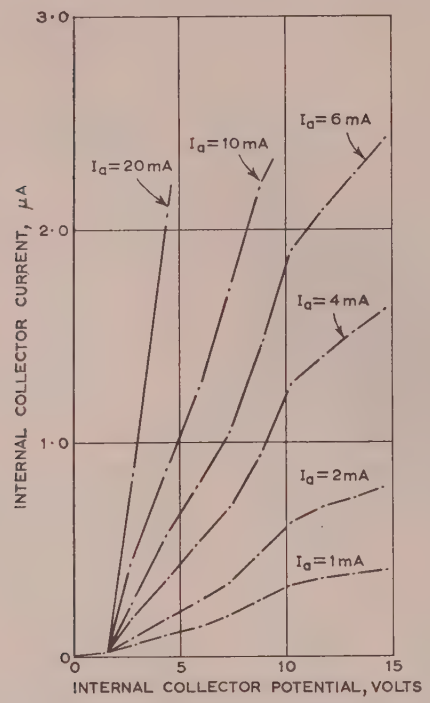


Fig. 6.—I_f/V_a characteristics for the experimental valve.

constant value $k_0 I_a$, which will be approximately 0·002 μA for $V_a = 250$ volts. With increasing positive values of V_g the value of I_f grows rapidly and depends on the magnitude of I_a. Except for very small values of I_a, it will be apparent that I_f will always be sufficient to swamp either or both of the radiation currents, $k_0 I_a$ and I_p.

(7.2) Course taken by Glass during X-ray Irradiation

Although X-ray irradiation is a potentially powerful means of raising the glass potential, it has not in practice been found to lead to lock-up conditions. The reason for this apparent anomally is sought in terms of the discussion in Section 7.1. As soon as anode current flows an emission of magnitude $k_0 I_a$ starts to charge the glass system. The existence of I_a, however, automatically implies the existence of I_f, and any rise of glass potential is immediately encountered by a compensating low-energy electron bombardment. The potential rise due to X-ray effect is therefore self-extinguishing.

(7.3) Effect of I_f on the Photo-electric Effect

Measurements of the rise of glass potential under conditions of photo-electric emission were taken with I_a biased back to zero on the control grid. If a milliampere or more is allowed to pass to the anode during the measurement, the presence of I_f clamps the glass potential to within 2–3 volts of earth potential. An interesting case arises when an illuminated valve is used for electronic switching. As soon as the control grid is made sufficiently negative to reduce I_a to zero (or a few microamperes) the glass potential begins to rise, and in some 1–30 sec reaches a steady value dependent on anode voltage and earth-leakage resistance, R_g. If this value is such that $\delta < 1$, a resumption of

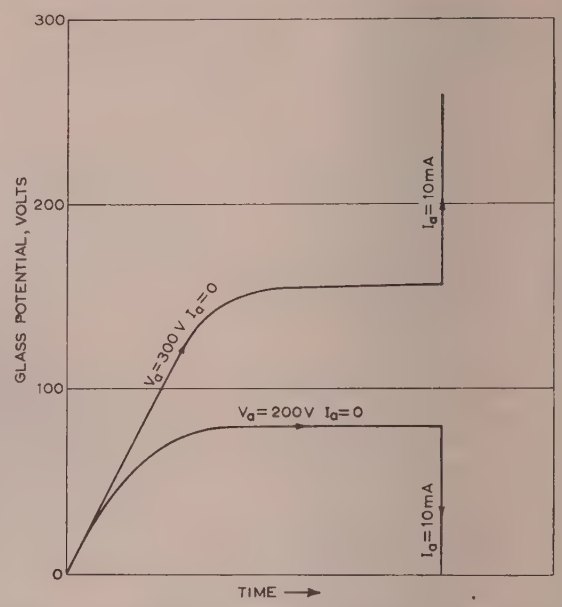


Fig. 7.—Grid-switching experiment at anode potentials of 200 and 300 volts.

anode current brings the glass potential back to earth; if $\delta > 1$ a lock-up condition occurs. The phenomenon is shown in Fig. 7 for two values of anode potential corresponding to $\delta < 1$ and $\delta > 1$. Once the lock-up condition is established it appears to be quite stable and maintains itself until broken down by the reduction of V_a to zero.

(8) CONCLUSION AND ACKNOWLEDGMENT

The general conclusion drawn from these experiments is that a lock-up condition can develop only if the glass potential has built up to such a value that $\delta > 1$ before anode current starts to flow. The three principal causes of build-up are anode-voltage leakage, capacitive transfer of anode voltage and photo-electric emission; the latter may be of interest in connection with electronic switching systems.

Acknowledgment is made to the Engineer-in-Chief of the General Post Office and to the Controller of Her Majesty's Stationery Office for permission to publish the paper.

(9) REFERENCES

(1) Benjamin, M., Cosgrove, C. W., and Warren, G. W.: 'Modern Receiving Valves: Design and Manufacture.' *Journal I.E.E.*, 1937, **80**, p. 401.
(2) McKay, K. G.: 'Advances in Electronics' (Academic Press, New York, 1948), Vol. 1, p. 97.
(3) Koller, L. R.: 'Secondary Emission, Part 1,' *General Electric Review*, April, 1948, **51**, p. 33.
(4) Deketh, J.: 'Fundamentals of Radio Valve Technique' (Philips Gloeilampenfabrieken, Eindhoven, 1949), Chapter 31.
(5) Metson, G. H.: 'Vacuum Factor of the Oxide-Cathode Valve,' *British Journal of Applied Physics*, 1950, **1**, p. 73.
(6) Metson, G. H.: 'The Physical Basis of the Residual Vacuum Characteristic of a Thermionic Valve,' *ibid.*, 1951, **2**, p. 46.

621.317.74 : 621.372.8 : 621.385.16.029.6

Paper No. 2037 R
May 1956

AN X-BAND MAGNETRON Q-MEASURING APPARATUS

By J. R. G. TWISLETON, B.Sc., Associate Member.

(*The paper was first received* 14*th October*, 1955, *and in revised form* 5*th January*, 1956.)

SUMMARY

The Q-factors of a resonant-cavity magnetron are usually estimated, in terms of the frequency bandwidth of a resonance curve, from measurement of the standing wave which exists in a feeder through which power is supplied to the magnetron at, and near to, its resonant frequency. The accuracy of this method is high, but when changes in Q-factors are to be measured which, for example, may result from changes in the output coupling or tuning of the magnetron, its application is tedious.

The use of a directional coupler replacing the standing-wave indicator to monitor the reflected power is well known. Such a coupler may be used in conjunction with a frequency-swept oscillator and cathode-ray tube to give a visual display of the magnetron resonance curve. The paper describes an apparatus of this type which was used for the qualitative examination of magnetron resonances in the 3 cm waveband, and details of an improved equipment which may be used to measure magnetron Q-factors are given, with a discussion of the accuracy of measurement which may be achieved.

LIST OF SYMBOLS

r = Voltage reflection coefficient due to the magnetron.
Q_x, Q_0 = External and internal Q-factors of the magnetron, respectively.
f_0 = Resonant frequency of the magnetron.
λ_0 = Wavelength corresponding to f_0.
δf = Frequency deviation from f_0, i.e. $f - f_0$.
$\delta\lambda$ = Deviation in wavelength corresponding to δf.
D, C = Directivity and coupling factors, respectively, of directional coupler.
K_1, K_2, K_3 = Alternative coupling factors for directional coupler.

(1) INTRODUCTION

In the construction of experimental magnetrons, it is usually necessary to determine the external and internal Q-factors of each valve prior to the final assembly, to ensure that the external coupling and circuit efficiency of the completed valve will be within the desired limits.

The Q-factors of a cavity magnetron can be deduced from the behaviour of the standing wave which exists in a feeder through which power is supplied to the magnetron. The method[1,2,3] comprises the measurement of voltage standing-wave ratio and the position of the standing wave at frequencies near to the resonant frequency. It is normally carried out using a travelling probe with a crystal detector. The Q-factors are calculated from the bandwidth of the resonance curve, but it is also necessary, in order to distinguish between the external and the internal Q-factors, to determine the way in which the position of the standing wave varies at frequencies near to the resonant frequency.

Only four travelling-probe measurements need be made, namely the measurement of the maximum and minimum voltage in the standing wave and the position of the standing wave at the resonant frequency and at an adjacent frequency. But, in practice, the resonant frequency and v.s.w.r. at resonance are usually found from a series of measurements made at frequencies near to the resonant frequency.

This method has the advantage of high accuracy, but when changes in Q-factors are to be measured resulting from, for example, changes in output coupling, or tuning of the magnetron, a large number of measurements are necessary and the method is very tedious.

The use of a directional coupler to monitor the reflected power is well known,[4] and it may be used, in conjunction with a frequency-swept oscillator and cathode-ray tube, to give a display of the magnetron resonance curve. An apparatus of this type is described, which was used for the qualitative observation of magnetron resonances in the 3 cm waveband. Details of an improved equipment which may be used to measure the magnetron Q-factors are given, with a discussion of the accuracy of measurement which may be achieved.

(2) DISPLAY OF THE MAGNETRON RESONANCE CURVE

In the basic apparatus, power is fed from a klystron, type 723A/B, through a directional coupler to the magnetron. A fraction of the power reflected from the magnetron appears in an opposite arm of the coupler, which is terminated by a crystal receiver, and an output signal is obtained proportional to the power reflected from the magnetron.

The klystron is frequency-modulated over a range of approximately 50 Mc/s by applying a 50 c/s sine-wave voltage to the reflector electrode, and a cathode-ray-tube display of the magnetron resonance curve is obtained with the power-output characteristic of the klystron superimposed upon it.

To evaluate magnetron Q-factors from the display of the Q-factor curve, scales of reflection coefficient (or v.s.w.r.) and frequency are required. Since the electronic tuning characteristics of the klystron vary with the nominal setting of frequency, it is difficult to devise a geometric scale which can be used under a variable set of conditions.

For this reason, use of the apparatus was confined to the qualitative examination of magnetron resonances. An improved apparatus will be described which enables the reflection coefficient to be measured, and hence for magnetron Q-factors to be determined.

(3) DESCRIPTION OF THE Q-MEASURING EQUIPMENT

The layout of the equipment is shown in Fig. 1. Two directional couplers, A and B, are connected in series. Coupler A is used to sample the power in the forward wave from the klystron, which is frequency-swept by the method previously described, and coupler B monitors the power reflected by the magnetron as in the original apparatus. The output 4 of coupler A and the output 3 of coupler B are each terminated with a crystal receiver preceded by a precision variable attenuator. The crystals in these terminations are chosen to have the same law, i.e. variation of output voltage with input power, and also approximately the same sensitivity.

Two rectified signals are obtained from the directional couplers, and these are selected alternatively by a mechanical switch

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. Twisleton is with the British Thomson-Houston Co., Ltd.

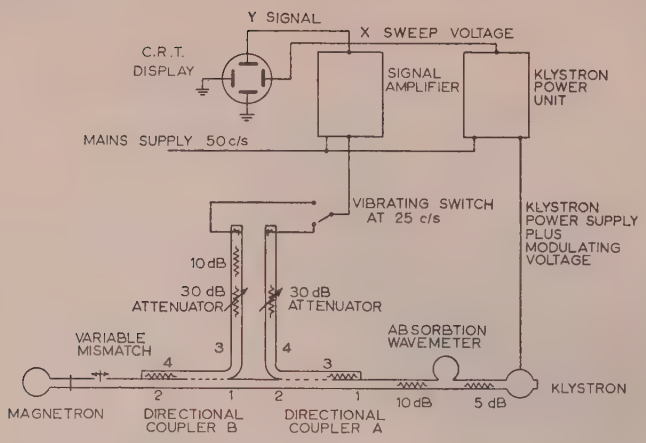


Fig. 1.—Schematic of magnetron Q-measuring apparatus.

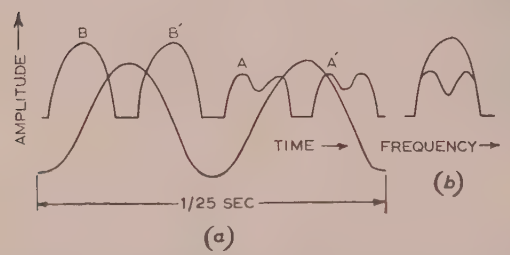


Fig. 2.—Voltage waveforms and displayed traces for 2-coupler system.
(*a*) Waveform of input voltage to amplifier.
(*b*) Displayed trace resulting from superposition of waveforms A and B.

(Carpenter relay) driven synchronously with the sweep voltage at 25 c/s and fed into the display unit [see Fig. 2(*a*)].

The trace of the signal from coupler B corresponds to the Q-factor curve of the magnetron and that from coupler A to constant reflection coefficient; the latter is used as a calibrating trace and its amplitude is controlled by the precision attenuator in arm 4 of coupler A. The two traces appear to be displayed simultaneously as shown in Fig. 2(*b*), and therefore, using the calibrating trace, the relative amplitudes of points on the magnetron Q-factor curve may be measured directly.

It is convenient to measure amplitude relative to unit reflection coefficient, and the apparatus is set up by substituting a movable short-circuit in place of the magnetron. With the short-circuit in position the attenuators are adjusted until the two traces are coincident. The magnetron is then replaced and the precision attenuator in arm 4 of coupler A is adjusted until the calibrating trace is coincident with that representing reflection from the magnetron at any particular frequency. The magnitude of the reflection coefficient may be calculated from the change in attenuation, x decibels, from the short-circuit condition:

$$r = 10^{-x/20} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

(4) MEASUREMENT OF MAGNETRON Q-FACTORS

The Q-factors of the magnetron are evaluated from the behaviour of the magnitude and phase of the voltage reflection coefficient as a function of frequency near to the resonant frequency.

The procedure involves the measurement of the reflection coefficient at resonance r_0 (or v.s.w.r.), and the determination of the width of the resonance curve, in terms of frequency, at a reflection coefficient r_1 which depends upon r_0. It may be shown that

$$r = \sqrt{\left[\frac{\left(1-\frac{Q_x}{Q_0}\right)^2 + \left(2\frac{\delta f}{f_0}Q_x\right)^2}{\left(1+\frac{Q_x}{Q_0}\right)^2 + \left(2\frac{\delta f}{f_0}Q_x\right)^2}\right]} \quad . \quad . \quad (2)$$

If a frequency deviation δf_1 is defined such that

$$2\frac{\delta f_1}{f_0} = \frac{1}{Q_x} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

the reflection coefficient r_1 at this frequency deviation is

$$r_1 = \sqrt{\left[\frac{\left(1-\frac{Q_x}{Q_0}\right)^2 + 1}{\left(1+\frac{Q_x}{Q_0}\right)^2 + 1}\right]} \quad . \quad . \quad . \quad (4)$$

The reflection coefficient at resonance is

$$r_0 = \left|\frac{1-\frac{Q_x}{Q_0}}{1+\frac{Q_x}{Q_0}}\right|$$

therefore, if Q_x is less than Q_0,

$$\frac{Q_x}{Q_0} = \frac{1}{\text{v.s.w.r. at resonance}}$$

and if Q_x is greater than Q_0,

$$\frac{Q_x}{Q_0} = \text{v.s.w.r. at resonance}$$

By finding the frequency deviation corresponding to $r = r_1$ from the resonance curve (as shown in Fig. 3), the external Q-factor may be calculated from eqn. (3). It should be noted

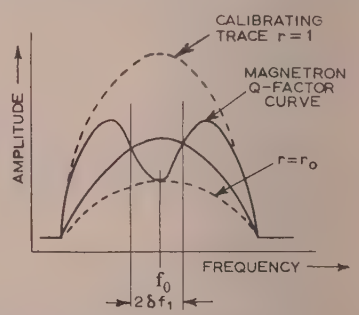


Fig. 3.—Displayed magnetron resonance curve.

that the Q-factors can be estimated from the width of the resonance curve measured at any height.[5] This general method is useful when r_0 is large, since it may then be desirable to measure the width of the resonance curve at a smaller height than r_1.

The question whether Q_x is less or greater than Q_0 may be decided from the variation in phase of the reflected wave at frequencies near to the resonant frequency; it cannot be deduced from the resonance curve alone.

(4.1) Measurement procedure for determining Magnetron Q-factors

The reflection coefficient at resonance is measured by setting the calibration curve to touch the resonance curve at its mini-

num point, as shown in Fig. 3; r_0 is then determined from eqn. (1).

Knowing r_0 the trace is set corresponding to the value of r_1 from eqn. (4), when it intersects the resonance curve at two points $\pm\delta f_1$. If Q_x is greater than Q_0 it may be shown that the Q-factor determined by the above method is Q_0.

Then either Q_x or Q_0 is equal to

$$\frac{f_0}{2\delta f_1} = \frac{\lambda_0}{2\delta\lambda_1}$$

The width of the resonance curve between these points is measured by means of a high-Q-factor absorption wavemeter inserted between the klystron and the directional coupler A. This wavemeter produces a reduction in the power fed into the waveguide system at its resonant frequency, which appears as a sharp dip in the displayed traces. The frequency at the point of cross-over of the traces may be found by tuning the wavemeter until the bases of the two 'pips' coincide, and two such measurements will give the width of the resonance curve.

The case of $Q_x < Q_0$ can be distinguished from $Q_x > Q_0$ by the behaviour of the resonance curve as a function of the phase of a mismatch relative to the magnetron. A variable mismatch, which may consist of a movable probe having a variable penetration into the waveguide, is inserted between the magnetron and the Q-measuring apparatus. As the distance of the mismatch probe, l, from the magnetron is varied, the observed resonant frequency and Q-factors vary cyclically. If the v.s.w.r. due to the mismatch alone, S_m, is less than the v.s.w.r. of the magnetron at resonance, S_r, the locus of the minimum reflection coefficient, which is a closed curve, appears as shown in Fig. 4(*a*). If S_m is

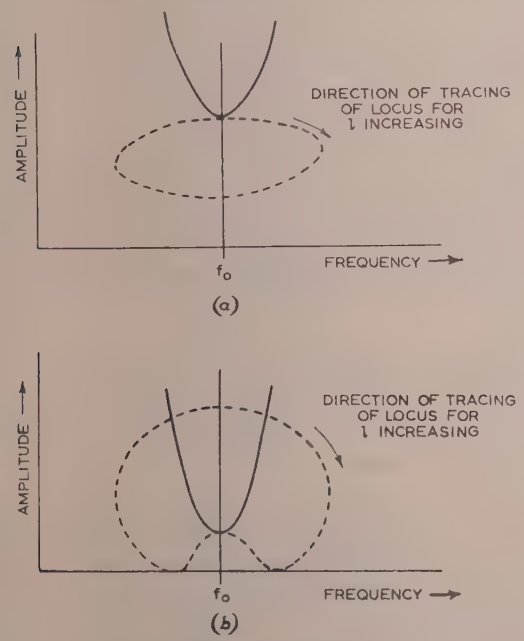


Fig. 4.—Behaviour of observed resonance curve with distance between magnetron and mismatch.
(*a*) $S_m < S_r$ and $Q_x < Q_0$.
(*b*) $S_m > S_r$ and $Q_x < Q_0$.

greater than S_r the locus of the minimum value of the reflection coefficient assumes a more complicated shape and touches the frequency axis at two points, one below and one above the resonant frequency of the magnetron.

The undercoupled case, $Q_x < Q_0$, and the overcoupled case, $Q_x > Q_0$, can be differentiated by observing the direction in which the locus of the minimum value of r is traced out as the distance, l, is varied. If, as l is increased, the rotation is clockwise, as shown in Fig. 4, Q_x is less than Q_0, and if the rotation is anti-clockwise, Q_x is greater than Q_0.

The direction in which the locus is traced out as S_m is varied remains unchanged, but a transition between the two forms of loci occurs when $S_m = S_r$.

The range of Q-factors which may be measured is limited by the sweep range of the klystron and also by the reflected power at resonance. As the reflected power at resonance increases, the variation in reflected power near to the resonant frequency decreases, and the voltage reflection coefficient becomes more and more indistinguishable from unity.

This inherent limitation in the method is also emphasized by the fall in the power output of the klystron at the ends of the sweep range. For these reasons the maximum value of Q_0 which may be measured, for $Q_x = 200$, is approximately 1 000.

(5) PERFORMANCE OF THE Q-MEASURING EQUIPMENT

(5.1) Measurement Errors arising in the Waveguide System

Errors arise in the measurement of reflected power using a directional coupler, if the 'directivity' of the coupler is imperfect. The 'directivity' and 'coupling' factors are usually defined as the ratio of powers obtained with all the arms matched, as shown in Fig. 5.

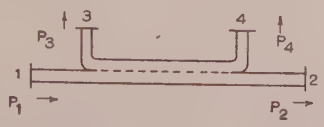


Fig. 5.—Power relations for directional coupler.

$$\text{Directivity} = \frac{P_3}{P_4} = \frac{K_2}{1 - K_1}$$

$$\text{Coupling} = \frac{P_4}{P_2} = \frac{K_1}{1 - K_1}$$

With all the arms correctly terminated

$$P_2 = (1 - K_1)P_1, \quad P_3 = K_2P_1, \quad P_4 = K_1P_1$$

Imperfect directivity introduces a phase dependency into the measurement of reflected power, which appears as a variation in the output voltage from the coupler with phase of the reflected voltage wave; the error may be large when the reflected power is comparable with the indirectly-coupled power.

In the 2-coupler system, the amplitude of the calibrating signal also depends to a small extent upon the phase of the reflected wave. A proportion of the power reflected by the magnetron appears at the crystal receiver of the coupler, which monitors the forward wave and produces an amplitude variation in the calibrating signal with phase of the reflected wave. Both effects combine to produce a relative amplitude variation between the calibrating signal and the signal which is proportional to the power reflected by the magnetron.

An estimate of this amplitude variation can be obtained for a given reflected power, in terms of the constants of the directional coupler. This enables the performance of the coupler to be specified in terms of the minimum accuracy of measurement which is allowable.

From Fig. 5 the power output from each arm is defined as a proportion K_1 of the input power. If P_1 is the input power, then, when all the arms are correctly terminated,

$$P_2 = (1 - K_1)P_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad (5),$$

and

$$P_3 = 0$$

If the termination of arm 3 produces a reflected wave, the power output from arm 3 is

$$P_3 = K_1(1 - K_1)P_1 r_2^2 \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where r_2 is the magnitude of the voltage reflection coefficient of the termination of arm 2, and the reflected power at the input arm of the coupler is $P_1(1 - K_1)^2 r_2^2$. Thus the voltage appearing at the termination of arm 3 is

$$V_3 = r_2 V_1 \sqrt{[K_1(1 - K_1)]} \quad . \quad . \quad . \quad . \quad (7)$$

If the directivity of the coupler is finite, the power output from the arm 3 is small but not zero in the matched condition. Then we may put $P_3 = K_2 P_1$, where $K_2 < K_1$. Because P_3 is small the distribution of power between the arms, as given by eqn. (6), will be almost unaltered so that, by superposition of voltages, the output voltage V_3 will now be

$$V_3 = V_1[\sqrt{K_2} + \sqrt{[K_1(1 - K_1)]} r_2 \varepsilon^{j\phi}] \quad . \quad . \quad (8)$$

The phase factor ϕ is arbitrary and indicates the relative phase difference between the voltage wave $V_1\sqrt{K_2}$ and the wave V_3 given by eqn. (7).

The fractional variation in V_3 for extreme values of phase is

$$\frac{\Delta V_3}{V_3} = \pm \frac{1}{r_2} \sqrt{\left[\frac{K_2}{K_1(1 - K_1)}\right]} \quad . \quad . \quad . \quad . \quad (9)$$

Thus, when $r_2 = \sqrt{[K_2/K_1(1 - K_1)]}$ the amplitude variation in V_3 may be $\pm 100\%$, depending upon the phase of r_2. If we put

$$\sqrt{\left[\frac{K_2}{K_1(1 - K_1)}\right]} = R \quad . \quad . \quad . \quad . \quad (10)$$

R is the voltage reflection coefficient at which the error in the measured reflected power may be $\pm 100\%$. This maximum error decreases according to the reciprocal law of eqn. (9), and the minimum error is $\pm R$, which occurs when r_2 is unity. Since it is impossible to distinguish, without separate measurements, whether r_2 is greater or less than R in the range

$$2R > r_2 > 0$$

it is important to achieve a low value of R.

If D and C are the directivity and coupling factors as defined by eqn. (1),

$$R = \sqrt{\left[\frac{D(1 + C)}{C}\right]} \quad . \quad . \quad . \quad . \quad . \quad (11)$$

Thus, for a directivity of -40 dB and a coupling of -7 dB, the minimum value of r_2 which may be measured to a degree of accuracy better than $\pm 10\%$ is 0·2.

A further source of error in the measurement of r_2 arises if the termination opposite the crystal detector monitoring the reflected power is reflecting. If the voltage reflection coefficient of this termination is r_4 the equation for V_3 becomes

$$V_3 = V_1\{\sqrt{K_2} + \sqrt{[K_1(1 - K_1)]} r_2 \varepsilon^{j\phi} + \sqrt{[K_1(1 - K_1)]} r_4 \varepsilon^{j\psi} \ldots\} \quad . \quad (12)$$

ψ is the phase of the wave which arrives in arm 3 due to reflection in arm 4. Thus the contribution due to r_4 increases or decreases R. It is preferable to attain a very low value for r_4 by using a matched termination when measurements are to be made over a wide frequency range.

In the 2-coupler system, an amplitude variation of the calibrating signal occurs owing to coupling between arms 2 and 4 of coupler A. The magnitude of this variation may be estimated in the manner previously described, and it is found that the fractional voltage variation in the calibrating signal V_4, for extreme phases of r_2, is

$$\frac{\Delta V_4}{V_4} = \pm r_2 \sqrt{\left[\frac{K_2(1 - K_1)^3}{K_1}\right]} = \pm r_2 \sqrt{\left[\frac{D}{C(1 + C)^3}\right]} \quad (13)$$

Thus the perturbation of the calibrating signal is proportional to r_2, and has the maximum value of

$$\pm \sqrt{\frac{D}{C(1 + C)^3}}$$

This is equal to $\pm R/(1 + C)^2$ if identical couplers are used, and for $D = -45$ dB and $C = -7$ dB, the maximum variation in the amplitude of the calibrating signal would be $\pm 0{\cdot}9\%$. In this case, the error due to perturbation of the calibrating trace is small compared with that arising in the direct measurement of reflected power by the coupler B.

(5.2) Characteristics of the Directional Couplers

It has been shown in Section 4.1 that the accuracy of measurement of reflection coefficient which may be achieved depends almost entirely upon the directivity of the couplers.

The couplers used in the apparatus were made to a design proposed by Riblet.[6] Each coupler has a cascade of 18 directive elements, which consist individually of two narrow slots cut at right angles to each other in the common broad face of the waveguide; the internal dimensions of the waveguide were 0·9 × 0·4 in. Within the wavelength range of 3·1–3·5 cm the directivity was better than 40 dB, and the coupling was approximately 7 dB.

The variation in the monitored reflected power with phase of the reflected wave was measured experimentally for one of the couplers using a calibrated mismatch, and the law of eqn. (9) was found to hold with $R = 0{\cdot}01$. From eqn. (11) the directivity was found to be -48 dB, assuming the coupling to be -7 dB.

(5.3) Performance of the Display Unit

Error arises in measurement of reflected power if the signals are distorted in the display system. Distortion, noise and interference can be minimized in the amplifier by suitable design, and in the apparatus described the error in the measurement of reflection coefficient is principally due to imperfect amplitude/phase discrimination of the reflected wave by the directional couplers.

(6) SUMMARY OF PERFORMANCE

The equipment may be used to measure the voltage reflection coefficient, r, of the magnetron, over a sweep range of 50 Mc/s within the 3 cm waveband. The maximum error in the measured value of r is less than $\pm 7\%$ for $r > 0{\cdot}3$ but the error increases as r decreases.

The range of Q-factors which may be measured is limited by the sweep range of the klystron and also by the magnitude of the power reflected at resonance. Resonances having a ratio of Q_x/Q_0 or Q_0/Q_x greater than 6 cannot be easily detected because of the small reduction in reflected power at resonance.

The range of measurement is sufficiently large to enable the Q-factors of most types of X-band magnetrons to be determined with sufficient accuracy and at greater speed than by the conventional method of plotting the resonance curve.

(7) ACKNOWLEDGMENTS

The investigation was carried out under a contract placed by the Department of Physical Research, Admiralty, and thanks are due to the Admiralty for permission to publish the paper. The

uthor is also indebted to his colleagues at the British-Thomson Houston Research Laboratory for encouragement and assistance n this work, particularly Mr. T. H. B. Baker.

(8) REFERENCES

(1) Latham, R., King, A. H., and Rushforth, L.: 'The Magnetron' (Chapman and Hall, 1952).
(2) Collins, G. B., Ed.: 'Microwave Magnetrons' (McGraw-Hill, 1948).
(3) Slater, J. C.: 'Microwave Electronics' (Van Nostrand, 1952).
(4) Southworth, G. C.: 'Waveguide Transmission' (Van Nostrand, 1950).
(5) Harmer, J. D.: 'Nomogram for Q of a Cavity', *Wireless Engineer*, 1955, **32**, p. 25.
(6) Riblet, H. J., and Saad, T. S.: 'A New Type of Directional Coupler', *Proceedings of the Institute of Radio Engineers*, 1948, **36**, p. 61.

DISCUSSION ON 'TRANSISTOR POWER AMPLIFIERS'*

MERSEY AND NORTH WALES CENTRE, AT LIVERPOOL, 5TH DECEMBER, 1955

Mr. G. Forshaw: The main consideration from the transistor application point of view is that the circuits and theory are still n advance of the actual device, at least the practical device available to users, but we hope to have a perfect practical ransistor very soon. As this will have to replace devices which have been in operation for a considerable time, it is essential hat it be reliable. The present limitations appear to be in the sealing and avoidance of contamination. Perhaps the authors will make some remarks on the developments of transistors as a reliable device.

The authors mentioned the limitations of temperature, 55° C o 75° C: I think the variations of parameters in that temperature range could have been amplified. We can provide temperature stabilization for an amplifier, but this further complicates the calculations considered in the paper.

At the beginning of Section 6.3 (common collector operation), he assumption is made that R_L is small compared with $r_c(1 - \alpha)$ as a generalization. To what extent is this assumption true? For transistors of small power-handling capacity it would appear permissible, but as the power rating is increased both α and r_c would appear to fall. For example, in a particular type of ransistor, $\alpha = 0{\cdot}96$ and $r_c = 20$ kilohms, as opposed to megohm, so that $r_c(1 - \alpha) = 800$ ohms, which, for $R_L =$ 00 ohms as quoted, would not permit of the approximation.

It would, then, appear to be possible to reduce R_L, but how ar can this reduction proceed without affecting the gain of the mplifier?

In non-sinusoidal applications the value of power in $R_L = 10P_{c\,max}$ would appear to be a value determined by the ype of transistor used. For example, with $R_L = 1{\cdot}2$ kilohms P.O. 3000-type relay), $V_c = 24$ volts and $I_{c\,max} = 20$ mA, witching rapidly for equal intervals, gives a mean load power f 240 mW and a mean collector power of 2 mW.

i.e. *On* 0·1 volt × 20 mA = 2 mW bottomed
Off 20 volts × 100μA = 2 mW saturated

iving a ratio P_{load}/P_C(mean) = 120.

There would also appear to be a dependency on the drive ondition, i.e. reduction of voltage V_{ce} with increased drive until stable condition is reached. Does the 2 millisec switching time eferred to in Section 9 apply to the relay times?

Would not the measure of P_{out}/P_{in} (mean) be a better measure f switching efficiency?

Mr. N. W. Morgalla: Eqn. 39 shows the dependence of power gain on $[\alpha_0/(1 - \alpha_0)]^2$, i.e. $(\alpha_0')^2$. This is a quadratic dependence. Disregarding frequency response, owing to large variations in this parameter from transistor to transistor (e.g. in one type, α_0 varies from 10 to 25), the variation of power gain could be approximately from 1 to 6 using different transistors, or a variation of 8 dB on the quoted 26 dB gain for common-emitter push-pull nominal value. Would this be one of the main practical disadvantages to this mode of working?

Secondly, in the non-sinusoidal application of relay operation, it is assumed that $V_{c\,max}$ is as stated in a previous Section—when I_{co} softens rapidly. What effect does the back-e.m.f. from the relay coil on release have on the transistor, since there is no mention of surge suppression being applied?

Mr. S. Beddoe: The gain of a thermionic valve amplifier can be increased by adding stages until a point is reached when shot effect and other valve noises become excessive. Do equivalent limitations exist for crystal devices? If so, what is the order of gain possible at audio frequencies on the basis of such limitations?

Messrs. R. A. Hilbourne and D. D. Jones (*in reply*): Mr. Forshaw has raised the question of reliability of junction transistors. With their poor sealing the early types were often lacking in this respect. However, owing to improvements in the manufacturing techniques the present types are very reliable. The general question of reliability is outside the scope of the paper.

The temperature limitation of 55° C–75° C mentioned is governed, not by the reversible variation of the transistor parameters, such as the variation of I_{co}, but rather by the irreversible variations of these parameters when the transistor is operated at elevated temperatures for long periods. The most serious of these is the deterioration of the current gain factor α_{cb}. Bias stabilization does not prevent these long-term variations.

As regards the lower output resistance of higher-power transistors, the particular transistor mentioned is intended to give about 2 watts of audio output power from a 6 volt h.t. supply. This necessitates a load resistance, R_L, of 9 ohms, which is considerably less than the 800-ohm output resistance quoted.

In general, the power limitation of transistors used to operate relays is mainly dependent on the power dissipated in the device while changing from one state to the other. The power dissipated in the transistor in the fully 'on' and fully 'off' conditions is very small, as was shown by Mr. Forshaw. The value of $P_{out} = 10P_{c\,max}$ applies to a change-over time which is approximately one-tenth of the time the relay remains in the 'end' state.

In reply to Mr. Morgalla, it is true that a large voltage pulse is developed across the collector junction when a transistor used for relay operation is switched off; this can be many times greater than the collector supply voltage and may produce large collector currents if the 'breakdown' voltage of the collector is exceeded. However, this pulse can be removed by connecting a junction diode across the relay coil as a voltage 'clamp.'

* Hilbourne, R. A., and Jones, D. D.: Paper No. 1861 R, April, 1955 (see 102 B, . 763).

Mr. Morgalla also raised the question of the power gain of a common-emitter amplifier varying as the square of the current gain factor, α_{cb}. The effect of this can be reduced by the application of negative feedback. It is also likely that manufacturers will tighten the range of α_{cb}; at present this is about 2 : 1 for most devices.

In reply to Mr. Beddoe, the ultimate signal that can be handled by a transistor amplifier is limited by noise. The noise factor of the amplifier, which is generally referred to some particular value of source resistance at the input, depends on a number of factors. Thus, at low frequency the noise power decreases inversely with increase of frequency; this variation occurs up to a frequency of a few kilocycles per second, the noise power at higher frequencies being substantially proportional to bandwidth. The noise factor decreases with decreasing emitter current but does vary appreciably with supply voltage. The minimum signal level that can be handled does, therefore, depend on a number of factors.

DISCUSSION ON 'AN 8-MeV LINEAR ACCELERATOR FOR X-RAY THERAPY'*

NORTH-EASTERN RADIO AND MEASUREMENTS GROUP, AT NEWCASTLE UPON TYNE, 7TH NOVEMBER, 1955

Dr. F. T. Farmer: I would like to make a few remarks based on experience with our own 4 MeV machine which has been in use at the Newcastle General Hospital for two years.

Primarily I would stress the thoroughly practical nature of these accelerators for generating X-rays and their value compared with some other types of megavolt equipment for therapy. The high output (100r/min or more) enables a greater number of patients to be treated per day than on conventional 250 kV apparatus, and the high percentage depth-dose makes radical treatment possible where other types of set fail. It appears that the principal advantages of super-high-voltage therapy are all realized very nearly to their full extent at the 4 MeV level. These are: high penetration in the body; absence of skin reaction through build-up in the first 8 mm or so; small degree of scatter and consequently minimal volume dose.

In addition, at this energy, the 'gantry mounting' is a practicable proposition with its important advantages for setting up patients. It would seem that this voltage, therefore, has some considerable merit in the field of megavoltage equipment.

Furthermore, since the linear accelerator is built in the main from standard radar components and does not require very specialized parts, the problem of replacements is less serious than in alternative high-voltage systems, and repairs can usually be carried out without great interruption of service. Magnetron lives are now 300–1 000 hours and are a less serious problem than was at first thought.

Dr. J. F. Fowler: Since the electron attains effectively the velocity of light within the first section of the accelerator waveguide, it would appear possible to build accelerators to give electrons of extremely high final energy, merely by adding identical later sections of waveguide. What, in fact, are the design problems involved?

Dr. B. C. Robinson: The author has mentioned the use of the 8 MeV and 4 MeV tubes for industrial and medical radiography. Could he give a brief indication of the respective fields covered by the various types of X-ray tube and of radio-active-isotope radiography.

Mr. C. W. Miller (*in reply*): Dr. Farmer's remarks on the clinical advantages of high-energy radiation and the ability of the linear accelerator to satisfy the demands of the radiotherapist are of great interest. His opinions carry considerable weight in view of his wide experience in the field. The equipment at Newcastle was made by an organization other than that with which I am associated, but it is basically similar to that mentioned in Section 7, and though having many engineering differences, was designed to meet the same specification. It would appear, therefore, that convenience and reliability of operation may be expected from any well-designed linear accelerator and are not only associated with one particular design.

Dr. Fowler is quite correct in his assumption that the linear acceleration of electrons may be continued to extremely high energies. A general account of the subject has been given elsewhere,* but it can be stated briefly that the design problems are those of providing sufficient radio-frequency power in correct phase relationship to the various sections of the accelerator. Limitations are eventually set by the difficulty in maintaining a large number of high-power valves, and, of course, economics also play a part. As regards very-high-energy machines, 600 MeV has already been achieved at Stanford University and an extension is contemplated. We have recently been occupied with the design of a 25 MeV high-power accelerator, which is now under construction. It is expected that the equipment will produce an electron beam with mean power of 30 kW.

Dr. Robinson's question is not easily answered in a few words. An account of the linear accelerator for industrial radiography has been published.† With an accelerator it is possible to provide radiation of higher energy, and therefore of greater penetrating power, than can be obtained from radioactive isotopes. Also, the accelerator can provide an intensity considerably greater than that from even very large radioactive sources. Thus, in the radiography of thick specimens, the exposure times using an accelerator are very much shorter than if isotopes were used. For example, using a 4 MeV linear accelerator, it is possible to radiograph a steel specimen 12 in thick in an exposure of one hour, while the exposure for the same sample using a 20-curie source of cobalt 60 would be approximately six months.‡

In medical radiography the chief advantage of high-energy radiation would appear to be reduced scatter and the relatively low absorption in bone. It becomes possible to identify details in a high-energy radiograph which would have been obscured in a more normal radiograph. Both in medical and in industrial radiography, the ability to focus the electron beam to produce a small focal spot, and thus produce a dimensionally small source of radiation, has great advantage in producing high resolution in the radiograph.

* MILLER, C. W.: Paper No. 1619, March, 1954 (see **101**, Part I, p. 207).

* MILLER, C. W.: *Engineering*, 1955, **180**, pp. 340 and 374.
† MILLER, C. W.: *Journal of the British Institution of Radio Engineers*, 1954, **14**, p. 361.
‡ MILLER, C. W.: *Engineering*, 1955, **180**, p. 455.

20.179.18 : 621.386.8 : 778.33

Paper No. 1881 U
July 1955

INTENSIFICATION OF THE X-RAY IMAGE IN INDUSTRIAL RADIOLOGY

A. NEMET, Dr.Sc.Techn., F.Inst.P., Member, and W. F. COX, B.Sc.

The paper was first received 23rd November, 1954, and in final form 21st April, 1955. It was published in July, 1955, and was read before the UTILIZATION SECTION *15th December, 1955, and the* NORTH-WESTERN UTILIZATION GROUP *31st January, 1956.)*

SUMMARY

The paper deals with the application of the X-ray image-intensifier tube in industrial fluoroscopy and radiography. It considers the factors governing resolution—i.e. brightness, blurring and contrast—and discusses the improvement expected from the use of the image intensifier.

Industrial fluoroscopy with this technique is compared with conventional fluoroscopy and contact radiography. Results are given of measurements obtained with D.I.N. penetrameters for steel and aluminium in the voltage range 30–200 kV.

The conclusion drawn is that the image intensifier allows new applications of industrial fluoroscopy and miniature radiography owing to the enhanced resolution which brings it near to contact radiography. It therefore permits more economical X-ray testing.

(1) INTRODUCTION

Since the discovery of X-rays there has been a steady increase in detail perceptibility, screen brightness and radiographic speed. However, the need for further improvement is realized in all types of X-ray examination.

In medical diagnosis the reduction of the X-ray dose received by the patients is very desirable, whereas in industrial radiology this problem does not arise, but the economy of the examination is important. Progress has been achieved by the development of fluorescent screens, films and the specific loading of the X-ray-tube focus.

The sensitivity of fluoroscopy was much inferior to that of radiography, however, until the introduction of the image intensifier, which also enabled patient dosage to be reduced.

The image intensifier dealt with by the authors was developed originally for medical diagnosis, and the paper describes its use in industrial radiology and shows that in this field also it enables an improvement to be made in sensitivity and economy.

(1.1) Description of the Tube

The image-intensifier tube used for the work recorded in the paper has already been described.[1] Fig. 1 shows the cross-section of the tube, the diameter of the X-ray screen being 5 in and the electron-image reduction ratio, 9 : 1.

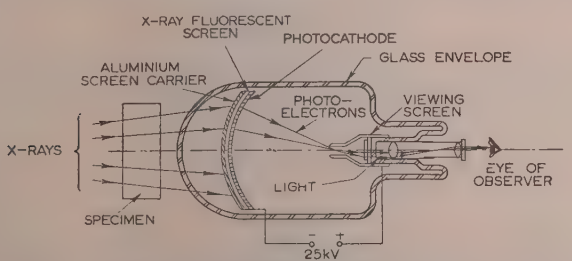


Fig. 1.—Sectional sketch demonstrating the principle of the image-intensifier tube.

The viewing screen is observed by a microscope during fluoroscopy, or is photographed in radiography and cine-radiography. The total brightness intensification is about 1 000 : 1, of which the image reduction produces a factor of 80. The remaining intensification—between 12 and 15—is due to acceleration of the photo-electrons modified by conversion processes in the two screens.

The intensifier fitted with a binocular microscope, together with its power unit, is shown in Fig. 2.

(1.2) Statistical Fluctuations

The effect of the quantum nature of X-rays is shown very strikingly by the tube. X-rays are transmitted in discreet quanta distributed at random, the number of which falling on a small area of a fluorescent screen during the accumulation time of the eye (0·2 sec) determines the brightness of that point. Owing to the random nature of quanta, this number—and hence the brightness—will vary according to the law of probability. The effect occurs in all stages of the fluoroscopic process and is most important in the stage where the number of quanta involved is least. This has been examined and described in two papers[1,2] in which it is shown that, in normal fluoroscopy, the important stage is the absorption of light quanta in the retina, whereas in image-intensifier fluoroscopy it is the absorption of X-ray quanta in the X-ray screen.

With high brightness intensification this random nature of the X-ray intensity may be directly demonstrated as a fluctuation in the contrast pattern.

The quantum fluctuations set a fundamental limit to the detail that can be seen in a fluoroscopic image. The smallest perceptible contrast produced by X-ray absorption in the object is about three times the relative standard deviation of the number of quanta producing the fluctuations.[3] This therefore controls the lower limit of the X-ray intensity, irrespective of the intensification.

The high brightness gain, however, allows fluoroscopy to be carried out in almost normal room lighting and with reduced dark adaptation time: it is therefore of economic value. The intensifier is so designed that part of the potential gain is used to reduce screen blurring, the overall blurring due to all causes being 0·3 mm referred to the X-ray screen.

For image-intensifier radiography, using f/1·5 lenses (see Section 2.1) and fast mass miniature radiographic films such as Flurodak and Fluorapid, the fluctuations are slightly worse than for image-intensifier fluoroscopy at normal brightness. With the same lenses and a slow fine-grain high-contrast copying film (Microfile) the number of quanta involved is about ten times greater than in image-intensifier fluoroscopy.

(2) FUNDAMENTAL CONSIDERATIONS

The perceptibility of a detail is governed by brightness, contrast and blurring. These three factors are not independent and should be considered together in a common theory to make it possible to evaluate the quality of a complete fluoroscopic or radiographic arrangement. Such a complete treatment, however, would need data not at present available, would be rather

Dr. Nemet and Mr. Cox are at Philips Balham Works Ltd.

Fig . 2.—Image-intensifier tube and power unit. The tube is fitted with a binocular microscope.

complex and would probably be only of limited value owing to the difficulties and restrictions of practical radiology. The three factors will be treated separately, and consideration will be given to the way in which they can be improved by the image intensifier. In Sections 3 and 4 the same effects will be discussed experimentally.

(2.1) Brightness and Exposure

The ability of the eye to detect small details and low contrasts improves with increasing brightness.[4] Rose[5] has shown that this improvement is due primarily to reduction of statistical fluctuations. The brightness level of normal fluoroscopy is very low, being of the order of 1 000 or more times lower than that at which a radiograph is viewed. An increase in fluoroscopic brightness will improve detail perceptibility, but this must be considered in conjunction with the quantum statistics as discussed in Section 1.2. The authors would, however, expect a considerable improvement in fluoroscopy with the image intensifier based on measurement of quanta in connection with the medical application of the tube.[1,3]

In radiography, brightness does not directly affect the image quality, but long exposure times are uneconomical. Indirectly the exposure determines the choice of screens, and therefore blurring, and also the tube voltage which controls contrast.

The viewing screen of the image intensifier can be conveniently photographed on 35 mm film, using two normal camera lenses and magnification of about unity. With lenses of f/1·5 aperture about 6% of the light emitted by the viewing screen reaches the film, whereas in normal radiography with intensifying screens nearly all the light is utilized. In spite of this low optical efficiency, image intensifier radiography is very fast. Flurodak and Fluorapid films have, at a density of one, approximately 20 to 40 times the speed, and Microfile the same speed, as normal radiography with high-definition calcium-tungstate screens.

The high speed of the fast films can be used to increase the specimen thickness, other factors being kept constant. For example, the authors have found that the increases in thickness of steel relative to radiography with high-definition calcium tungstate screens are 0·20 in and 0·55 in at 100 kV and 160 kV, respectively, with Flurodak brand of film, and 0·24 in and 0·68 in at the same voltages with Fluorapid.

(2.2) Contrast

Contrast in fluoroscopy is defined as the relative change of brightness, $\Delta I/I$, and if this is produced by a cavity of thickness Δx in a material of linear absorption coefficient μ,

then
$$\frac{\Delta I}{I} = \frac{\mu \Delta x}{1 + I_S/I_D}$$

where I_S/I_D is the ratio of brightness due to scattered and direct radiation. The ratio I_S/I_D can have values of 14 or more,[6] and depends on the X-ray beam diameter and the absorber thickness, and generally increases with voltage in the range under consideration.[7] Since I_S/I_D increases and μ decreases with increasing voltage, a low h.t. supply should be used to obtain a high contrast. In normal fluoroscopy this is impossible since the X-ray-tube voltage must be made very high—in the range of 50–80 kV above that for normal radiography—in order to obtain sufficient brightness.

Using the image intensifier for fluoroscopy, sufficient brightness and a large gain in contrast can be obtained at relatively low voltages. With decreasing voltage, however, the statistical fluctuations will increase, and an optimum value may be expected for which perceptibility is a maximum. A loss of contrast in the image-intensifier tube must also be mentioned, which is only about 10% if the image is of fairly uniform brightness. If there are large brightness differences in the image, the contrast reduction will be much greater where the brightness is low.

In radiography, the contrast is higher than in fluoroscopy by the film gradation γ, other factors affecting contrast being equal. The values of γ for normal radiography are 2–3 for lead screens or no screens, and 3–4 for tungstate screens.[8] In image-intensifier radiography, using normal X-ray developer, $\gamma = 3$ for the Microfile film, whereas $\gamma \simeq 1{\cdot}7$ with Flurodak and Fluorapid fast films.

(2.3) Blurring

The summation of the three components of blurring—i.e. the geometrical blurring (B_g) due to the finite size of the focus, the screen blurring (B_s) and the movement blurring (B_m)—has been discussed in relation to medical radiography and fluoroscopy.[9,10,11]

For industrial radiology we can usually exclude movement blurring and we can add the geometrical and screen blurrings by the simple formula,*

$$B = \sqrt{(B_s^2 + B_g^2)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

proposed by Newell.[9]

It has been shown[12,13] that, using X-ray magnification, the effect of the total blurring can be reduced. If f is the focus size, B_e the effective blurring and M the magnification,

then $$B_g = f(M-1) \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

and $$B_e = \frac{B}{M} = \frac{\sqrt{[B_s^2 + f^2(M-1)^2]}}{M} \quad . \quad . \quad . \quad (3)$$

The effective blurring has a minimum with respect to M when,

$$M_{opt} = 1 + \left(\frac{B_s}{f}\right)^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

and the corresponding effective blurring is given by

$$B_{e\,min} = \frac{B_s}{\sqrt{\left(1 + \frac{B_s^2}{f^2}\right)}} = \frac{f}{\sqrt{\left(1 + \frac{f^2}{B_s^2}\right)}} \quad . \quad . \quad . \quad (5)$$

These formulae apply directly to normal fluoroscopy and radiography and image-intensifier fluoroscopy. In image-intensifier radiography there is, in addition, blurring due to the lens and the film, which is equivalent to a further blurring B_f in the X-ray screen of the image intensifier. B_f is the sum of the film blurring and the optical blurring in the image on the film multiplied by the reduction factor of the image intensifier. In image-intensifier radiography, B_s in eqns. (1), (3), (4) and (5) should thus be replaced by B_c, the sum of B_s and B, and the authors have assumed

$$B_c = \sqrt{(B_s^2 + B_f^2)}$$

The value of B_s for the screens available for normal fluoroscopy is about 0·7 mm and for the image intensifier 0·3 mm. The blurrings B_c for calcium-tungstate high-definition and high-speed intensifying screens[14] are 0·2 and 0·35 mm. For normal radiography with lead screens or no screens we may assume a value of $B_c = 0{\cdot}05$ mm.[15] The authors have measured the combined optical and film blurring for the three films used with the image intensifier and the resolution in the film image was found to be about 50 lines/mm for Microfile film and 30 lines/mm for both Flurodak and Fluorapid films. This is equivalent to blurrings of 0·18 mm and 0·30 mm respectively at the X-ray screen, and the corresponding values of B_c are 0·35 mm and 0·42 mm.

The minimum effective blurring and corresponding optimum magnification for several fluoroscopic and radiographic techniques have been calculated using the above equations. Fig. 3(*a*)

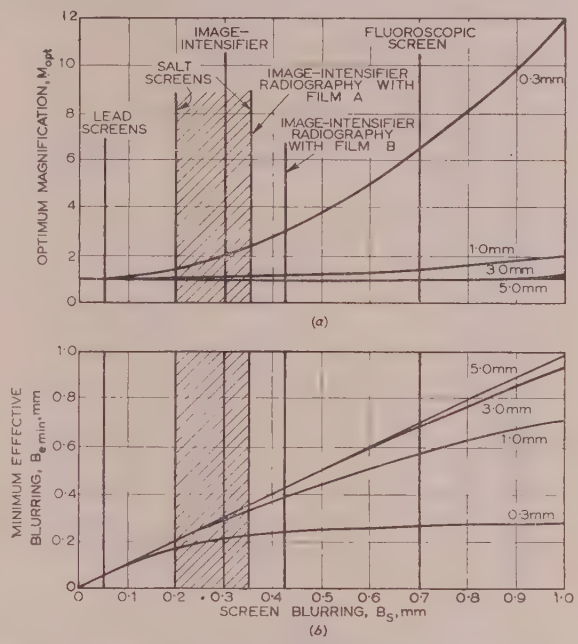


Fig. 3.—Optimum magnification and minimum effective blurring as functions of screen blurring.

(*a*) Optimum magnification as function of screen blurring.
Film A: Fine-grain slow film (Microfile).
Film B: Fast film (Flurodak and Fluorapid).

(*b*) Minimum effective blurring as function of screen blurring.

shows the optimum magnification plotted as a function of the screen blurring with the focus size as parameter. Four commonly used focus sizes, 0·3, 1·0, 3·0 and 5·0 mm, have been chosen. In Fig. 3(*b*) the minimum effective blurring as a function of screen blurring is shown for the same focus sizes.

For normal fluoroscopy with a focus of 0·3 mm, $B_{e\,min}$ is 0·27 mm and M_{opt} is 6·5. With this large value of magnification the focus-screen distance can be reduced—until limited by the specimen size—with a consequent increase in screen brightness. With the 3 mm and 5 mm foci the magnification must be low, and the focus-screen distance consequently high, and the screen brightness low.

In image-intensifier fluoroscopy, the only useful improvement due to magnification is with the 0·3 mm focus, for which $B_{e\,min} = 0{\cdot}21$ mm and $M_{opt} = 2$.

For normal radiography with salt screens a significant improvement is again obtained only with the 0·3 mm focus, the screen blurrings of 0·2 mm and 0·35 mm being reduced to 0·17 mm and 0·23 mm for magnifications of 1·4 and 2·4. Radiography without screens is not improved with magnification.

In image-intensifier radiography a useful reduction in blurring is obtained with the 0·3 mm focus for the slow and fast films used by the authors. $B_{e\,min}$ is 0·23 mm and 0·25 mm and M_{opt} is 2·4 and 3·0 respectively; the corresponding values of B_c are 0·35 mm and 0·42 mm. Only a 10% blurring improvement

* According to Klasens, $B = (B_1^3 + B_2^3)^{1/3}$ is somewhat more accurate than eqn. (1), but the authors have used eqn. (1) for simplicity of calculation.

can be obtained with the 1·0 mm focus. Image-intensifier radiography has therefore the same blurring as radiography with high-definition salt screens, and, for slow Microfile film, almost the same contrast.

The values of magnification with the 0·3 mm focus for both fluoroscopy and radiography with the image intensifier are not critical and can be varied from about 1·6 to 3·5 with only a small change in B_e.

Fig. 4 shows the effective blurring B_e as a function of magnification. The results are for the same four foci as used previously, for two values of screen blurring and for one value of film blurring. It is often difficult to obtain magnifications less than 1·1, indicated by the shaded area in the Figure. With the aid of these curves it is possible to assess the blurring of a complete radiographic or fluoroscopic arrangement, and to decide the value of magnification.

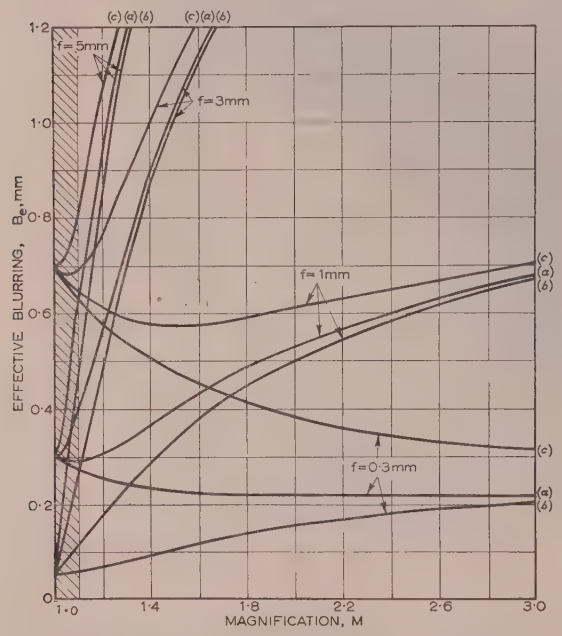


Fig. 4.—Variation of effective blurring with magnification.

(*a*) $B_s = 0·3$ mm.
(*b*) $B_s = 0·05$ mm.
(*c*) $B_s = 0·7$ mm.
$B_e = \sqrt{[B^2 + f^2(M-1)^2]}/M$.

(3) PENETRAMETER MEASUREMENTS

To obtain a direct measure of the overall gain due to the use of the image-intensifier tube, penetrameter measurements were carried out with two equipments. The first used a 0·3 mm rotating-anode X-ray tube and was limited to 125 kV; the second was a 200 kV equipment having a 5 mm stationary-anode tube. Both used half-wave generators; the first had two-valve rectification and the second was a self-rectifying tube-head unit.

Tests were made on aluminium and steel plates, and comparative measurements were made for normal fluoroscopy, image-intensifier fluoroscopy, contact radiography and image-intensifier radiography.

Penetrameters are designed to obtain an overall numerical assessment of a fluoroscopic or radiographic process in such a way that the ability to detect a flaw is tested direct. The three partial effects of brightness, blurring and contrast are thus assessed together. Owing to the diversity of the flaws that may have to be detected by radiological processes, the design of the penetrameter itself is not yet universally agreed, the two main types being (*a*) the German D.I.N. penetrameter using wires of different diameters, and (*b*) the step-wedge type of penetrameter which is often preferred in this country and in the United States. The sensitivities* obtained with these two types of penetrameter will generally differ. An interesting comparison of penetrameter types has been made by Curtis[16] with aluminium, using as test objects, spheres, wires and plates. He found that the sensitivity figures decreased in that order, but that the difference between wires and plates became larger with decreasing wire diameter. The authors have observed the same effect with wire and step-type penetrameters, as shown in Table 1. It appears that the wire-type penetrameter corresponds more closely than the step-wedge type to the flaws generally found in radiographic and fluoroscopic examination, and the D.I.N. type has therefore been chosen for the measurement of sensitivities. The penetrameter was always placed on the X-ray-tube side of the material being examined, for which position the image quality is worse. The measurements shown in the following Tables are mean values for five observers.

Table 1

COMPARISON OF D.I.N. AND STEP-WEDGE PENETRAMETER SENSITIVITIES FOR FLUOROSCOPY OF ALUMINIUM WITH THE IMAGE INTENSIFIER

Thickness	Voltage	D.I.N. sensitivity	Step-wedge sensitivity
in	kV	%	%
$\frac{1}{8}$	40	11	4·5
$\frac{1}{4}$	50	7·9	3·4
$\frac{3}{8}$	60	6·3	2·3
$\frac{1}{2}$	65	4·7	2·2
$\frac{3}{4}$	75	4·2	2·4
1	85	3·5	2·0
$1\frac{1}{4}$	90	2·5	2·1
$1\frac{1}{2}$	100	2·6	2·4

Focus-screen distance: 60 in.
X-ray tube current: 10 mA.
Focus: 5·0 mm.

(3.1) Fluoroscopy

Measurements were made with the image intensifier and the apparatus having the 0·3 mm-focus tube, operated at 2 mA and a focus-screen distance of 20 in. The tube was viewed in artificial lighting of average brightness. The maximum thickness that could be penetrated was $3\frac{1}{2}$ in of aluminium and $\frac{3}{8}$ in of steel (both at 121 kV). In darkness, the thickness ranges extended to 5 in of aluminium and $\frac{3}{4}$ in of steel (both at 125 kV).

It was found that perceptibility increased up to a magnification factor of about two. Magnification was then fixed at 2, the calculated effective blurring being 0·2 mm.

For normal fluoroscopy the apparatus with the 5·0 mm focus was used and a high-definition fluoroscopic screen in contact with the specimen. The tube current was 10 mA and the focus-screen distance 60 in. The screen was viewed in darkness after an adaptation time of about ten minutes.

This apparatus was also used with the image intensifier under the same conditions except that the image was viewed in artificial lighting. The test material was placed in contact with the intensifier giving magnifications not exceeding 12% for objects up to 4 in thick. For this thickness the geometrical blurring was calculated to be 0·6 mm and the corresponding total blurring

* "Sensitivity" is the diameter of the thinnest penetrameter wire or step just visible divided by the thickness of the material and expressed as a percentage.

)·62 mm. The limits of the thickness of aluminium that could be viewed were 4 in in dim artificial lighting and 5½ in in darkness.

(3.1.1) **Aluminium.**

The test results for aluminium are given in Table 2A. The voltages shown give the best perceptibility except for the voltage limitation of the apparatus.

With the image intensifier and the 5 mm focus, sensitivities lie, as was expected, between the values obtained with normal fluoroscopy and the combination of the image intensifier with the 0·3 mm X-ray tube. The sensitivity figures are about twice those found with the 0·3 mm focus but considerably better than in normal fluoroscopy for all thicknesses of aluminium.

Table 2A

D.I.N. Penetrameter Sensitivities for Fluoroscopy of Aluminium with and without Image Intensifier

Thickness	Image-intensifier fluoroscopy						Normal fluoroscopy		
	0·3 mm focus (125 kV max) $M = 2$ Focus–screen distance: 20 in 2 mA			5·0 mm focus (200 kV max) $M = 1·08$ Focus–screen distance: 60 in 10 mA			5·0 mm focus (200 kV max) $M = 1·05$ Focus–screen distance: 60 in 10 mA		
	Voltage	Sensitivity	Wire diameter	Voltage	Sensitivity	Wire diameter	Voltage	Sensitivity	Wire diameter
in	kV	%	in	kV	%	in	kV	%	in
⅛	30	6·3	0·008	40	11	0·014	80	19	0·024*
¼	34	3·9	0·010	50	7·9	0·020	100	12·5	0·032*
½	48	2·4	0·012	65	4·7	0·024	100	7	0·040*
1	63	2·0	0·020	85	2·4	0·024	130	5·9	0·060
1½	77	1·6	0·024	100	2·6	0·040	160	5·3	0·080
2	88	1·4	0·028	120	2·3	0·048	180	5	0·100
2½	99	1·3	0·032	120	2·8	0·072	190	4·7	0·120
3	107	1·0	0·032	160	2·6	0·080			
3½	121	1·1	0·040	190	3·4	0·120			
4	125	1·2	0·048†	200	3·9	0·160			
4½	125	2·2	0·100†						
5	125	2·4	0·120†						

* Similar figures obtained with 0·3 mm focus at ×1 magnification.
† Measured in darkness.

Table 2B

D.I.N. Penetrameter Sensitivities for Fluoroscopy of Steel with and without the Image Intensifier

Thickness	Image-intensifier fluoroscopy						Normal fluoroscopy		
	0·3 mm focus (125 kV max) $M = 2$ Focus–screen distance: 20 in 2 mA			5·0 mm focus (200 kV max) $M = 1·08$ Focus–screen distance: 60 in 10 mA			5·0 mm focus (200 kV max) $M = 1·05$ Focus–screen distance: 60 in 10 mA		
	Voltage	Sensitivity	Wire diameter	Voltage	Sensitivity	Wire diameter	Voltage	Sensitivity	Wire diameter
in	kV	%	in	kV	%	in	kV	%	in
⅛	81	4·7	0·006	80	12·5	0·016	200	16	0·020
¼	103	3·9	0·010	105	7·8	0·020	200	11	0·028
⅜	121	3·7	0·014	130	7·3	0·028	200	9·5	0·036
½	125	3·2	0·016*	150	7·1	0·036	200	9·6	0·048
⅝	125	3·2	0·020*	—	—	—	—	—	—
¾	125	4·3	0·032*	200	5·3	0·040	200	8·5	0·064
1				200	4·8	0·048	200	8·0	0·080
1¼				200	6·4	0·080			
1½				200	7·0	0·105*			

* Measured in darkness.

Referring to the image intensifier and the 0·3 mm focus, one can see that the sensitivity figure initially improves with increasing thickness to 1% at 3 in, and deteriorates above 4 in where the brightness is decreasing. With normal fluoroscopy the sensitivity improves with thickness to 5%, which is about the same as other workers have found.[17,18] For a given thickness, the image-intensifier sensitivity is about three times better than in normal fluoroscopic technique. Between 1 in and 4 in the sensitivity is between 1% and 2% and is therefore comparable with radiography using fast salt screens.

(3.1.2) **Steel.**

Measurements for steel are shown in Table 2B.

These measurements show a similar dependence of sensitivity on thickness to the one for aluminium. Again, approximately the same improvement in sensitivity figures holds. It is interesting that 125 kV apparatus allows a useful fluoroscopic penetration of ¾ in steel.

(3.2) **Radiography**

Image-intensifier radiographs were made with a 35 mm camera having two f/1·5 miniature radiographic lenses of focal length

2 in. Microfile, Flurodak and Fluorapid films were used and all were developed in D.19b X-ray developer for 5 min at 68° F. Radiographs were made with the 0·3 mm-focus apparatus only, which was operated at 4 mA, an X-ray magnification of 2·0 and a focus-screen distance of 20 in. A constant exposure time of 40 sec was chosen for all cases, since this was considered a reasonable compromise between voltage reduction and exposure time. The measurements obtained with Microfile and Flurodak films produced some vignetting in conjunction with Microfile, but this was not noticeable with Flurodak.

(3.2.1) **Aluminium.**

The sensitivities with Flurodak are about the same throughout as they were for fluoroscopy, but for Microfile there is an improvement, increasing with thickness, from 1·2 to 2 times. The optimum voltage with Microfile was found to be about 40% of

Table 3A

D.I.N. Penetrameter Sensitivities for Radiography of Aluminium with and without the Image Intensifier

Thickness	Image-intensifier radiography 0·3 mm focus (125 kV max.) $M = 2$ Focus–screen distance: 20 in 4 mA. 40 sec exposure						Normal radiography					
	Slow film			Fast film			Fine-grain film			Non-screen film		
	Voltage	Sensitivity	Wire diameter	Voltage	Sensitivity	Wire diameter	Voltage	Sensitivity	Wire diameter	Voltage	Sensitivity	Wire diameter
in	kV	%	in	kV	%	in	kV	%	in	kV	%	in
$\frac{1}{4}$	28	3·2	0·008	—	—	—	60	1·56	0·0039	60	2·36	0·0059
$\frac{1}{2}$	32	2·0	0·010	28	2·8	0·014	75	1·18	0·0058	75	1·58	0·0079
$\frac{3}{4}$	35	1·6	0·012	31	2·1	0·016						
1	39	1·6	0·016	33	2·0	0·020	100	1·19	0·0119	100	1·57	0·0157
$1\frac{1}{2}$	46	1·3	0·020	39	1·3	0·020						
2	53	1·2	0·024	44	1·4	0·028	120	0·60	0·0119	120	0·79	0·0157
$2\frac{1}{2}$	58	1·0	0·024	48	1·3	0·032						
3	66	0·8	0·024	53	1·1	0·032						
4	80	0·6	0·024	61	1·0	0·040	220	0·59	0·0236	220	0·79	0·0315
5	98	0·6	0·032	71	0·8	0·040						
6	112	0·8	0·048	84	1·6	0·100						
7	125	0·9	0·064	96	2·0	0·140						

Table 3B

D.I.N. Penetrameter Sensitivities for Radiography of Steel with and without the Image Intensifier

Thickness	Image intensifier radiography 0·3 mm focus (125 kV max) $M = 2$ Focus–screen distance: 20 in 4 mA. 40 sec exposure						Normal radiography					
	Slow film			Fast film			Fine-grain film			Non-screen film		
	Voltage	Sensitivity	Wire diameter	Voltage	Sensitivity	Wire diameter	Voltage	Sensitivity	Wire diameter	Voltage	Sensitivity	Wire diameter
in	kV	%	in	kV	%	in	kV	%	in	kV	%	in
$\frac{1}{8}$	52	4·8	0·006	44	4·8	0·006	120	<3·12*	0·0039	120	<3·12*	0·0039
$\frac{1}{4}$	67	2·4	0·006	53	4·8	0·012	140	1·56	0·0039	140	2·36	0·0059
$\frac{3}{8}$	88	2·7	0·010	63	3·7	0·014						
$\frac{1}{2}$	101	2·0	0·010	72	2·8	0·014	180	1·24	0·0062	180	1·58	0·0079
$\frac{3}{4}$	123	1·6	0·012	94	2·6	0·020						
1				109	3·1	0·032	220	1·18	0·0118	220	1·18	0·0118
$1\frac{1}{4}$				122	2·5	0·032						

* Limited by end of penetrameter wire range.

are shown in Table 3; the sensitivities of the two fast films were found to be the same.* The films were projected so that the image was the same size as that on the X-ray screen, i.e. twice the size of the object. It was found, however, that projection down to about natural size made no apparent change in the detail that could be seen. The optical system of the camera that used in normal radiography, and with Flurodak approximately 35%. The Microfile sensitivies are generally less favourable than those of normal radiography, the ratio ranging between 1·3 and 2 for the fine-grain film and 1·3 to 1·5 for the ordinary film. In spite of the voltage differences the exposure times were approximately the same as in ordinary radiography, calculated for the same distances.

* No direct comparison of the blurrings of the two fast films was made.

It is noteworthy that aluminium thicknesses up to 7 in could be radiographed with 125 kV with a sensitivity of 0·9% and 40 sec exposure.

(3.2.2) **Steel.**

The improvement in sensitivity between Microfile radiography and image-intensifier fluoroscopy increases with thickness. The sensitivities with Microfile are slightly worse than for normal radiography. The Flurodak sensitivities are comparable with those found in fluoroscopy with the same apparatus.

The maximum thickness for Microfile was $\frac{3}{4}$ in and for Flurodak $1\frac{1}{4}$ in. Again the exposures for the image intensifier are about the same as for normal radiography, although the voltages are only about one-half.

It has been found that a 40 sec exposure for Flurodak demands great care to avoid afterglow, and it is likely that for practical work the exposure time of this film will have to be reduced by a factor of 3–4.

(4) PRACTICAL RESULTS

Castings and welds, mainly of aluminium and steel, were examined fluoroscopically and radiographically with the image-intensifier and conventional techniques.

For image-intensifier fluoroscopy the 0·3 mm-focus X-ray tube was operated at 2 mA and 20 in focus-screen distance with a magnification of 2. For normal fluoroscopy the same apparatus was operated with a 1·0 mm-focus X-ray tube at 5 mA, 20 in focus-screen distance and a high-definition screen with no magnification. Normal fluoroscopy was performed in darkness and image-intensifier fluoroscopy in artificial lighting.

When the object did not completely intercept the X-ray beam it was found both in normal and image-intensifier fluoroscopy that the high brightness caused by the unattenuated X-rays reduced visibility of detail in the image. This effect was more marked with the image intensifier owing to scatter in the tube and the higher brightness ratio between the surrounding field and the image due to the lower voltage. Some improvement was obtained by immersing the specimens in solutions of lead salts, but the best results were achieved by stopping the surrounding X-ray beam with a lead mask. This was used with conventional and image-intensifier fluoroscopy when necessary. The same effect occurs in image-intensifier radiography, and here again lead masks were used in a similar way. For best results it is therefore advisable to work with as small an X-ray beam as possible. It was found with the image intensifier that perceptibility did not critically depend on voltage over a range of about 30 kV.

The arrangements for image-intensifier radiography were exactly the same as for the radiographic penetrameter measurements.

In Sections 4.1 and 4.2, the results obtained with a few specimens in fluoroscopy and radiography, respectively, are described. Normal and image-intensifier radiographs of some of these objects are shown in Figs. 6, 7 and 8. Normal radiographs were made on a fine grain film without screens except where indicated.

It should be noted that some detail is lost in the reproduction, and that the density range of the normal radiographs is much reduced. Some spurious black spots appear in the image-intensifier radiographs owing to imperfections in the image-intensifier tube. The sizes of details quoted were measured on the contact radiographs, which have negligible blurring.

(4.1) Fluoroscopy

A cast-aluminium plate, 0·72 in thick and containing cavities ranging in size from about $\frac{1}{4}$ in downwards, was examined, and a contact radiograph was made with 0·002 in lead screens at 80 kV [Fig. 5(*a*)]. Image-intensifier fluoroscopy showed all the cavities

(*a*)

(*b*)

(*c*)

Fig. 5.—Radiographs of cast-aluminium plate 0·72 in thick containing cavities.

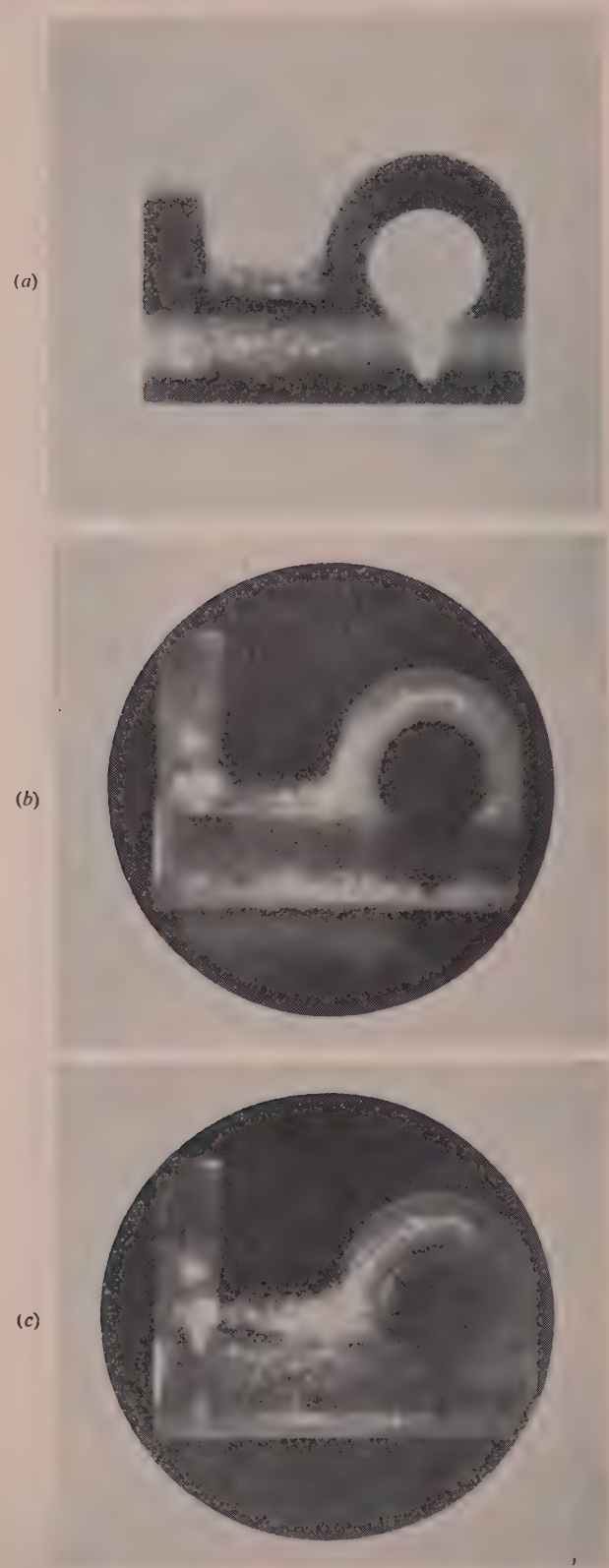


Fig. 6.—Radiographs of aluminium-bearing casting of irregular shape with sections extending to ¾ in thickness.

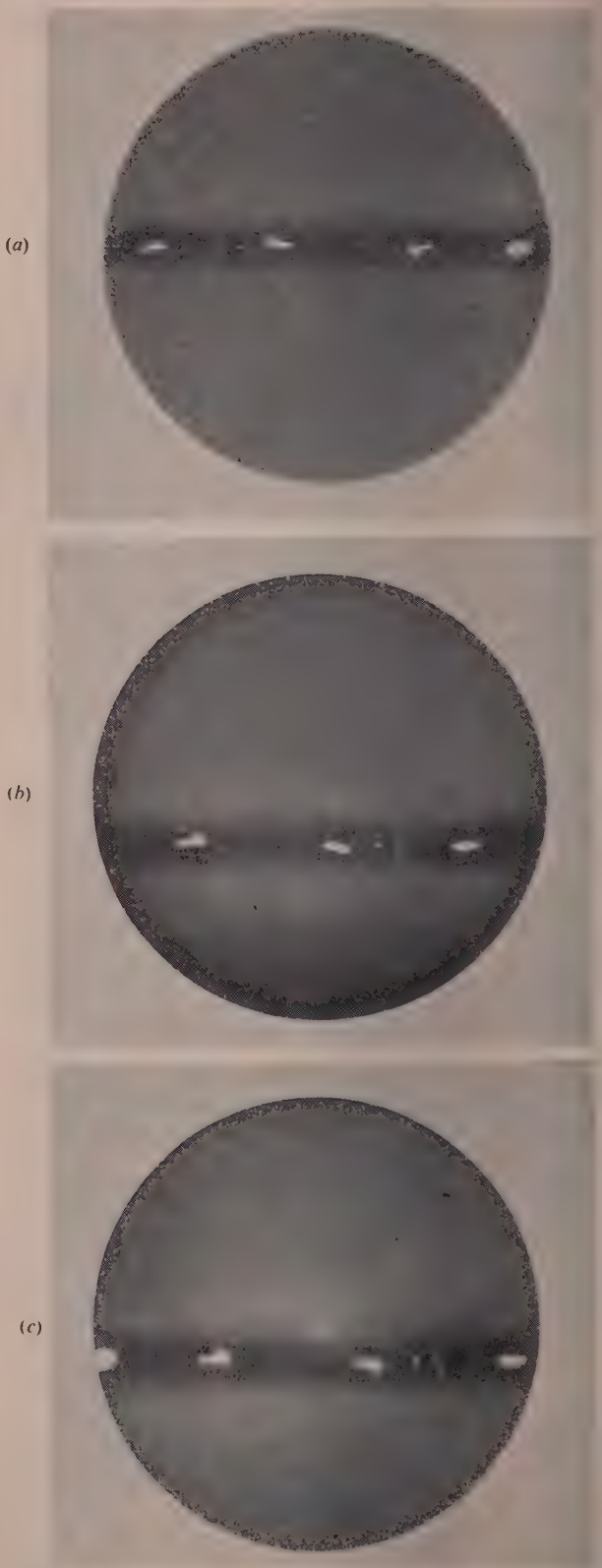


Fig. 7.—Radiographs of an arc weld of steel plates 0·15 in thick containing cavities.

down to about 0·035 in at a voltage of about 50 kV. Normal fluoroscopy at 125 kV showed the larger cavities only, the smallest being 0·080 in.

An aluminium bearing casting was examined, having sections extending in thickness up to $\frac{3}{4}$ in and showing extensive porosity on a contact radiograph made at 80 kV [Fig. 6(*a*)]. A large proportion of the porosity could be seen with the image intensifier. Diffuse spots, which in the contact radiograph were of about 0·030 in diameter, could be seen. The optimum voltage was about 60 kV. Ordinary fluoroscopy at 125 kV showed the porosity as a diffuse patch containing no detail.

A weld between steel plates 0·15 in thick contained cavities varying in size from 0·013 in upwards and also fine cracks. Fig. 7(*a*) is a contact radiograph of this specimen taken at 100 kV. Image-intensifier fluoroscopy at 70 kV showed all the cavities of diameter above about 0·024 in and gave a blurred image of the larger cracks without showing the detail seen in the contact radiograph. Ordinary fluoroscopy at 120 kV showed only the cavities larger than about 0·080 in, and none of the cracks.

A piece of $\frac{3}{8}$ in steel plate containing a range of fine surface cracks was also examined. A radiograph with lead screens at 120 kV showed the larger cracks and a number of the finer ones. Image-intensifier fluoroscopy at 120 kV showed all the larger cracks exceeding about 0·020 in width and 0·010 in depth. Normal fluoroscopy at 200 kV at 10 mA and 60 in focus-screen distance showed the largest cracks of about 0·040 in width and 0·020 in depth.

An aluminium–zinc alloy casting with a wall thickness of 0·1 in was radiographed at 100 kV, and was found to contain cavities with diameters above 0·005 in. Image-intensifier fluoroscopy at 70 kV showed defects with diameters larger than about 0·012 in. Normal screening at 125 kV showed the largest defects only with a minimum diameter of about 0·040 in.

The above results show that, using the image intensifier with a fine-focus tube, magnification technique and reduced voltage, a considerable improvement is obtained over normal fluoroscopy. The minimum size of cavities that can be detected is reduced by a factor of between $2\frac{1}{2}$ and 3 compared with normal screening. It was also found that defects can be detected much more easily and quickly with the image intensifier.

(4.2) Radiography

With the cast-aluminium plate 0·72 in thick the contact radiograph [Fig. 5(*a*)] showed cavities, the smallest one visible being of 0·008 in diameter. Radiographs with the intensifier were made with Flurodak and Microfile at 30 kV and 34 kV respectively, and are shown in Figs. 5(*b*) and 5(*c*). The smallest cavities displayed are of about 0·040 in and 0·024 in diameter for the two films, and are in each case about $2\frac{1}{2}$ times the diameter of the smallest D.I.N. wires visible under the same conditions. The contrast in the Microfile film is much greater than that in the Flurodak one and detail is more easily seen. Compared with the normal radiograph, the Flurodak radiograph has less contrast and much of the smaller detail is lost, whereas the Microfile radiograph appears to have about the same contrast and less detail is lost.

Fig. 6(*a*) is the normal radiograph of an aluminium bearing casting with sections up to $\frac{3}{4}$ in thick, and containing flaws varying in size down to about 0·010 in. This object had sections varying in thickness up to $\frac{3}{4}$ in, and in order to record all sections on one exposure with the intensifier it was embedded in a mixture of table salt and 5% red-lead powder which had the same X-ray absorption as aluminium. The voltage used for the Flurodak film [Fig. 6(*b*)] was 34 kV and for the Microfile film [Fig. 6(*c*)] 41 kV. The first film showed details of diameter larger than 0·040 in, and the latter one about 0·024 in. As with the first object, the higher contrast of the Microfile film was very marked and details in it could be more easily seen than in the Flurodak film.

Fig. 7(*a*) is a normal radiograph, made at 100 kV and without screens, of a weld between 0·150 in-thick steel plates. It shows cracks of widths between 0·004 in and 0·015 in, and cavities with a minimum size of 0·018 in, but smaller ones could have been resolved by the method.

The exposures for the image intensifier radiographs on Flurodak and Microfile films, Figs. 7(*b*) and 7(*c*), were made at 50 kV and 60 kV respectively. On the fast-film radiograph, cavities down to about 0·024 in are visible, and the group of fine cracks which are completely traversing the weld appear as a blurred streak without the individual cracks being distinguishable. The shorter cracks at the edge of the weld are not visible. With the Microfile film, cavities down to diameters of about 0·018 in are seen, as also are the short cracks at the edge of the weld.

Experimental work with image-intensifier radiography shows that Microfile film enables one to see smaller details than Flurodak film. In agreement with Sections 3.2.1 and 3.2.2, it appears that if the speed of the fast film is used to obtain a higher contrast by lowering the voltage, the advantage gained is insufficient to compensate for the high gamma of the slow film. The contrast of the slow-film radiographs appears always to be very similar to that of normal radiography with high-definition calcium-tungstate screens.

It is important to note, however, that the density range obtainable with miniature radiographs is less than that with contact radiographs, even if the film-screen combinations have the same gamma and blackening. The reason is that in projection the denser parts of the film cannot be penetrated. This is not, however, a serious disadvantage since it is offset by the fact that two or more radiographs of differing exposures of a given object can be quickly taken at little expense.

(5) CONCLUSIONS

(5.1) Fluoroscopy

From the theoretical considerations in Sections 2.1–2.3 the value of the image intensifier in industrial fluoroscopy can be summarized. It has been seen that the blurring of the image intensifier is about one-half of that of the normal fluorescent screen, and if a 0·3 mm focus is used and magnification of two, the effective blurring can be reduced to one-third of that in normal fluoroscopy.

The contrast is increased by using voltages lower than those used in normal radiography. In addition, visual acuity is increased and dark adaptation can be omitted or considerably reduced. A significant improvement in detail perceptibility can therefore be expected.

The penetrameter measurements show that these improvements are obtained in practice. The sensitivities for aluminium are in the range of 1–2%, and for steel 3–4%, with an accompanying valuable increase in thickness penetrated for a given voltage, as can be seen in Table 2.

It is therefore likely that problems can now be solved with the image intensifier which previously could not be dealt with either by fluoroscopy, on account of poor resolution, or by radiography, for economic reasons.

The reduction of voltage will reduce the size of fluoroscopic generators and ease the protection problem in the design of stands and handling equipment. Our practical results with the image intensifier were obtained with the X-ray tube and intensifier mounted in a cabinet of $1\frac{1}{2}$ mm steel lined with $\frac{1}{2}$ mm lead. This reduced the stray-radiation dose rate to below tolerance level up to 125 kV. Mass-production examination on conveyor belts of objects such as bearings, etc., would appear

to lend itself to this technique. For long welds, the handling equipment would have to be designed to move the plates along the image intensifier. There is no doubt a wide field for mechanical-engineering design which will have to suit the particular requirements rather than attempt to make universal handling equipment.

An example of an installation designed for fluoroscopy applied to the production of large-diameter steel pipes (up to 10 m in length) is shown in Fig. 8. The X-ray tube is mounted on a long stem and is stationary, so arranged that the tube window is opposite the image intensifier. This is mounted in an observation cabin having protective walls. The pipes to be examined pass on rails in front of the image intensifier and surround the X-ray tube. The horizontal weld to be examined is observed during the travel and the doubtful places are marked.

Fig. 8.—Fluoroscopy applied to the production of steel pipes.

The X-ray tube operates on 150 kV d.c., is water cooled and has a focal spot of 0·4 mm. The magnification used is 2 : 1, the distance between intensifier screen and tube focus being 20 cm. It is claimed that wall thicknesses up to 30 mm can be examined. Wire sensitivities of 2½–3% are being achieved. It is understood that up to 10 000 m of pipe-line have successfully been examined by this method by two manufacturing firms in Germany.

For research purposes and type testing in laboratories, and for foundries as an aid to developing casting techniques, a simple cubicle fitted with remote-handling tongs may be of general value.

(5.2) Radiography

In radiography with the image intensifier and Microfile film, both contrast and blurring may be expected to be comparable with normal radiography using salt screens. The fast film should give similar blurring with less contrast. This again is confirmed by penetrameter measurements and practical results. The principal gain is in the size and cost of the film, the latter being from 3 to 12 times less than normal film costs corresponding to image-intensifier magnifications 2:1. Further, there is the advantage of the reduced storage space for film records as compared with full-size films, which may be valuable for large-scale survey radiography such as in projectile check. The size of the image-intensifier screen will, however, limit the size of the specimen that can be radiographed with one exposure. In general, because of this, the method will be best applied to production checking of small articles. For irregular shaped objects which do not entirely intercept the X-ray beam, masking is essential to obtain good contrast.

Owing to the limited density range (0–1·5) that can be used in projection, the thickness range covered in an image-intensifier radiograph is less than that in a normal radiograph.

The most striking aspect of image-intensifier radiography is the voltage reduction for a given thickness. With the Microfile film, normal radiographic voltages can be halved. This should lead to considerable economy in the X-ray generator for a given task.

(6) ACKNOWLEDGMENTS

The authors wish to acknowledge the valuable help received from Mr. J. C. Rockley, of the Aeronautical Inspection Directorate of the Ministry of Supply, who kindly provided the penetrameter measurements for normal radiography. They are also indebted to Mr. R. Halmshaw and the late Mr. C. Croxson, of the Armament Research Establishment of the Ministry of Supply, for their interest in this work and for providing specimens. Mr. A. C. Carswell and Mr. R. W. Shaw of High Duty Alloys, Ltd., have also provided specimens.

Acknowledgment is due to Mannesmann Forschungsinstitut G.m.b.H., Duisburg-Huckingen, Phönix-Rheinrohr A.G., Mühlheim, and Dr. R. O. Schumacher of C. H. F. Müller A.G., Hamburg, for the data on pipe-line examination and

) the second-mentioned firm for the photograph reproduced n Fig. 8.

(7) REFERENCES

(1) TEVES, M. C., and TOL, T.: "Electronic Intensification of Fluoroscopic Images," *Philips Technical Review*, 1952, **14**, p. 33.

(2) STURM, R. E., and MORGAN, R. H.: "Screen Intensification Systems and their Limitations," *American Journal of Roentgenology and Radium Therapy*, 1949, **62**, p. 617.

(3) OOSTERKAMP, W. J., and TOL, T.: "Prinzipielle und praktische Grenzen der Detailerkennbarkeit bei verschiedenen röntgenologischen Beobachtungsmethoden insbesondere bei Verwendung der Bildverstärkerröhre," *Fortschritte auf dem Gebiete der Röngenstrahlen vereinigt mit Röntgenpraxis*, 1954, **81**, p. 381.

(4) MOON, P.: "The Scientific Basis of Illuminating Engineering" (McGraw-Hill Book Co., 1936), chapter 12.

(5) ROSE, A.: "The Sensitivity Performance of the Human Eye on an Absolute Scale," *Journal of the Optical Society of America*, 1948, **38**, p. 196.

(6) CROWTHER, J. A. (Editor): "Handbook of Industrial Radiology" (Arnold and Co., 1944), p. 16.

(7) NEMET, A., COX, W. F., and HILLS, T. H.: "The Contrast Problem in High Kilovoltage Medical Radiography," *British Journal of Radiology*, 1953, **26**, p. 185.

(8) CROWTHER, J. A. (Editor): "Handbook of Industrial Radiology" (Arnold and Co., 1944), p. 72.

(9) NEWELL, R. R.: "Sharpness of Shadows in Radiography of the Lungs," *Radiology*, 1938, **30**, p. 493.

(10) KLASENS, H. A.: "The Blurring of X-Ray Images," *Philips Technical Review*, 1947–48, **9**, p. 364.

(11) NEMET, A., COX, W. F., and WALKER, G. B.; "Blurring in Radiography," *British Journal of Radiology*, 1946, **19**, p. 257.

(12) BURGER, G. C. E., COMBEE, B., and VAN DER TUUK, J. H.: "X-Ray Fluoroscopy with Enlarged Image," *Philips Technical Review*, 1946, **8**, p. 321.

(13) O'CONNOR, D. T., and POLANSKÝ, D.: "Theoretical and Practical Sensitivity Limits in Fluoroscopy," *Non-Destructive Testing*, 1951–52, **10** (No. 2), p. 10.

(14) CROWTHER, J. A. (Editor): "Handbook of Industrial Radiology" (Arnold and Co., 1944), p. 82.

(15) HALMSHAW, R.: "The Factors involved in the Assessment of Radiographic Definition", *Journal of Photographic Science*, 1955, **3**, (to be published).

(16) CURTIS, L. R.: "Fluoroscopic Inspection of Aluminium and Magnesium Castings," *Non-Destructive Testing*, 1948–49, **7** (No. 3), p. 25, and 1949, **7** (No. 4), p. 24.

(17) MAYER, R. W.: "A Practical Comparison of Fluoroscopy and Radiography," *Industrial Radiography*, 1944–45, **3** (No. 3), p. 28.

(18) IONA, M.: "New Developments and Applications of Fluoroscopic Examination of Metals and Assemblies," *ibid.*, 1944, **2** (No. 4), p. 20.

DISCUSSION BEFORE THE UTILIZATION SECTION, 15TH DECEMBER, 1955

Dr. H. G. Taylor: Some three years ago the organization with which I am associated was invited to plan some non-destructive testing on what the Admiralty called 'stores'. At the time, we did not know anything about them, but were given to understand that the quantities were large and that the material was mild steel. We were not given any idea of their shape or other properties, but the rate of inspection was to be very high and it was most important that the welds should be satisfactory.

We thought hard about what we could do to ensure a high rate of non-destructive examination. Mr. Carson, a colleague of mine, worked on this problem, giving it a great deal of thought and considering many different methods—including ordinary radiographic methods. We considered the possibility of high-speed radiography, and wondered whether it was possible to use this in the normal way but at a really high speed, with fast development, wet examination of the negatives and so on. I then asked 'Why not use fluoroscopy?' and suggested that we might increase the voltage or take other steps to improve it.

Mr. Carson told me that there were objections to it, but he found the article quoted as Reference 1 in the paper, which deals with the image-intensifier tube produced in Holland. This seemed to be a complete answer to our problem, and we corresponded with the organization in question, but the result was nil because there was only one tube in existence at that time.

In the end, it was discovered that the rate of inspection was very much less than had at first been expected, so that ultrasonics were satisfactory; but it was an interesting exercise to be faced with the problem of high-speed inspection, which I feel sure occurs in industry, and probably in connection with things which have to be very accurately made, such as ammunition and fuses.

For cases of that kind the best solution seems to be fluoroscopy, using this image intensifier. The situation has entirely changed, for the engineer now has available a new tool with many applications.

One application which I had thought of, and had intended to put forward, is that which the authors showed in presenting the paper. I had in mind pressure vessels rather than pipes, but the application to pipes which was shown is even better, because in pipes, welds are not made to Class 1 standard, whereas for pressure vessels they usually are to this standard. I imagine that with this device it should be possible to combine the advantages of fluoroscopy and photography.

The inspection of pressure vessels by X-rays is very expensive. All pressure vessels of Class 1 standard have to be inspected completely, so it is necessary to take radiographs of the complete longitudinal and circumferential seams. The majority of the radiographs indicate that the weld is perfectly satisfactory, but all must be taken in order to find the odd occasion where the weld is defective and where the defect must be cut out and re-welding done.

It has been felt in the industry for many years that it is desirable to have a method of inspection which would find the defect quickly, and then take a radiograph of it. Is it possible with this image intensifier to inspect the welds by fluoroscopy and then, when a defect is seen, switch over or alter the optical arrangements and take a photograph there and then? If so, I think that in the welding field for the inspection of seams this apparatus has great possibilities.

Another practical application is the inspection of spot welds. There is one firm in this country which uses spot welding for aircraft primary structures. A large proportion of the welds are X-rayed. That is another case where it might be possible to examine the spot welds directly by fluoroscopy. This would be a much cheaper process than the present one.

I was going to ask the authors to say more about safety, but Mr. Cox has indicated that they have given a good deal of attention to the question of the risk of using the apparatus. I do not think that these details are in the paper, and it would be advantageous if they could be added, because the introduction of such new equipment will inevitably cause some concern from the point of view of health risk.

It is a pity that the authors do not give more details of the tube. A sketch is given in Section 1.1, and that Section is headed 'Description of the Tube', but it contains only five or six sentences, which is not adequate for an Institution paper. Does the accuracy of the result depend in any way on the perfection of the shape of the screen? It is curved for a purpose. Does it have to be optically shaped, or does that not matter very much?

The authors have kept a very level balance between fluoroscopy and radiography, but Mr. Cox, in presenting the paper, said he felt there was more scope for the device for fluoroscopy than for photography, and I fully agree. I think that is the most important aspect, and that it should have been brought out more clearly in the paper. It detracts from the value of the tube to balance so evenly the two aspects. So far as I can see, on the photographic side its advantages are not very great, but on the fluoroscopic side they seem to be very great.

Referring to photography, I had always thought that the observation of small defects in radiographed specimens was limited by the grain size of the film used. If that is the case, then using small film, to which the authors point as an advantage, seems to be distinctly a disadvantage, because presumably the grain size on small film is the same as on large, and the defect is so much smaller.

Mr. E. van Someren: I should like to take up the last point which Dr. Taylor made, because as a photographer I was interested in that problem. I have made enlargements from the small 35 mm films used on the image intensifier and compared them with my own radiographs of the same specimens, and I find that there is no loss of detail due to using those fine films.

I am also interested in the radiography of welds, and I feel that the demonstrations which I have seen of this apparatus show that it is suitable for the inspection of welds to what the Americans call the A.P.I.-A.S.M.E. standard, which is not quite up to our Class A pressure-vessel standard, in the range of $\frac{3}{8}$–$\frac{3}{4}$ in steel, which covers so much pipe welding. I would say, however, that if the job can be satisfactorily inspected with iridium 192 it is possible to get about the same amount of detail with image-intensifier fluoroscopy in thicknesses up to $\frac{3}{8}$ in of stainless steel or $\frac{1}{2}$ in of mild steel. I have had comparative radiographs made of a stainless steel weld with iridium 192 and by the image intensifier.

Direct fluoroscopic inspection gives an opportunity to detect a crack by twisting the weld while it is under observation so that the plane of the crack falls along the axis of the inspection beam. That is a very attractive 'talking point' for fluoroscopy, but, looking at the sort of pipes which have been inspected, one could not very well twist them during inspection to detect a crack. I wonder whether there has been any gain in the visibility of cracks under industrial conditions due to the ability to move the specimen while it is being inspected.

Mr. J. C. Rockley: The industrial application shown by Dr. Nemet—the examination of welded seams—is well suited to the instrument, since a single motion enables the whole weld to be covered completely. The examination of castings is a much more difficult problem, however, in view of the need to handle the specimen in such a way that the whole volume of the metal in the casting can be covered. Have the authors given any consideration to this aspect of the matter?

The penetrameter results are most impressive, and while they can be accepted as a guide to the sensitivity that can be obtained, it is equally important to consider the capabilities of the instrument in terms of different types of defects. To this end I have radiographed a few small light-alloy castings by normal film radiography and by radiography using Dr. Nemet's image intensifier. Fig. A illustrates microporosity in $\frac{3}{4}$ in thick magnesium alloy. This defect is invisible with normal screening

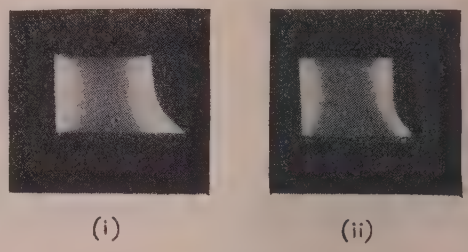


Fig. A.—Microporosity in $\frac{3}{4}$ in magnesium–aluminium alloy.
(i) Normal radiography.
(ii) Image intensifier.

methods. It is revealed by normal radiography, and again, but with greater contrast, by radiography with the image intensifier.

Fig. B shows pinhole porosity in an aluminium-alloy casting, and indicates how the area of coverage when using the image intensifier for radiography at $\times 2$ magnification is much reduced.

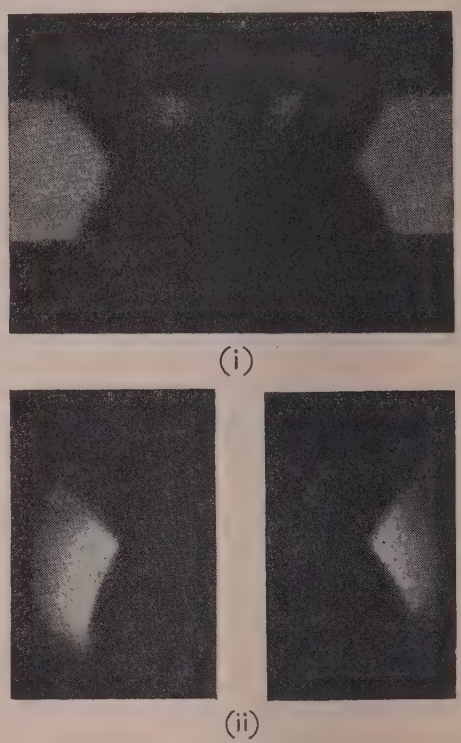


Fig. B.—Pinhole porosity in aluminium–silicon-alloy casting.
(i) Normal radiography.
(ii) Image intensifier.

The field covered is about $2\frac{1}{2}$ in in diameter, and thus necessitates a very large number of exposures to cover the area of a single exposure by normal radiography.

Fig. C illustrates a shrinkage crack in an aluminium–copper-alloy casting in which the original defect has become partially sealed by the copper-rich constituents of the alloy. This defect is faintly visible when using the image intensifier visually. This Figure clearly shows the enhanced contrast and reduced sharpness of the radiographs taken by the image intensifier. Fig. D shows another example of copper segregation. This example is barely visible by ordinary fluoroscopic methods, but is plainly revealed on the image intensifier. The two comparative radiographs, one

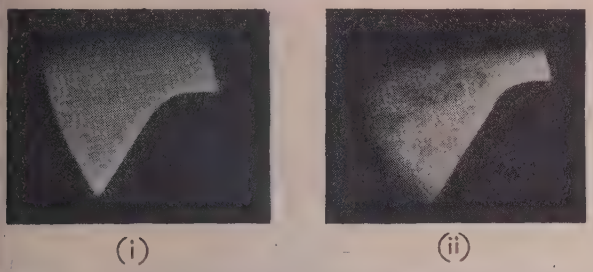


Fig. C.—Shrinkage crack in aluminium–copper-alloy casting.
(i) Normal radiography.
(ii) Image intensifier.

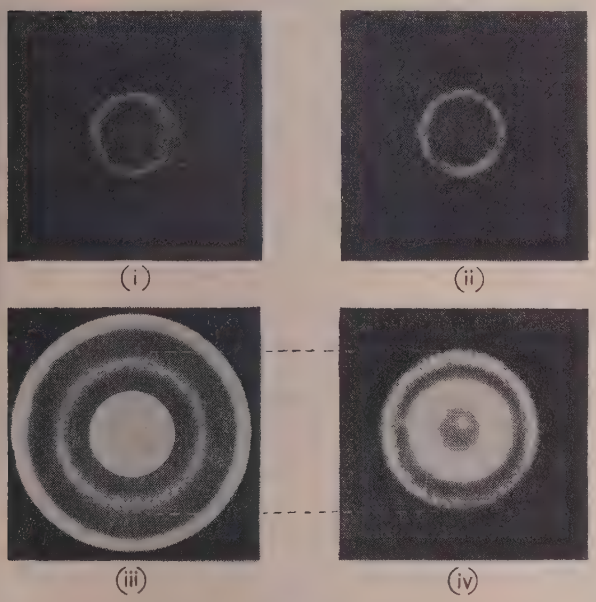


Fig. D.—Segregation in aluminium–copper-alloy castings.
(i) $\frac{1}{16}$ in thick: normal radiography.
(ii) $\frac{1}{16}$ in thick: image intensifier.
(iii) $\frac{1}{2}$ in thick: normal radiography.
(iv) $\frac{1}{2}$ in thick: image intensifier.

taken by normal radiographic means and the other by radiography on the image intensifier, again clearly indicate the increase in contrast and loss in definition in the latter case.

These are some of the examples on which we have had quite good results with this instrument. They are, however, restricted to thicknesses not exceeding 1 in, and it would be of interest to know whether the authors have any results to report on sections of greater thickness.

Mr. H. I. S. Allwood: The authors have referred to the scintillations observed as a result of the quantum nature of the X-ray beam. These arise in the process of absorbing the X-rays in the fluorescent screen in the first stage of the tube. The fact that they are visible in the final picture is an indication that there is not much more to be gained by further improvement in the later stages. In taking a single radiograph, in order to get a picture of the quality which is usually regarded as acceptable, there must be absorbed in the fluorescent screen several million X-ray quanta per square centimetre; otherwise the picture will be largely spoilt by these quantum fluctuations.

That being so, if we are to achieve any further reduction in X-ray exposure, we must ensure that a greater proportion of the X-ray quanta which emerge from the object under examination is absorbed in the fluorescent screen, or perhaps in some alternative X-ray-sensitive system. Can the authors say what proportion of the X-ray beam falling on the intensifier tube is usefully employed by being absorbed in the screen, and whether they have any hope that the proportion will be improved in the near future?

I should like to ask also whether the performance figures which have been quoted apply only to the centre of the image intensifier screen. At the low X-ray-tube kilovoltages which may be used the absorption and scattering in the hemispherical glass dome of the tube may vary appreciably across the screen. In addition, the electron optical system utilizes a curved cathode and a flat viewing screen. Do these effects in fact cause any loss of resolution in the picture towards the edge of the tube?

Mr. R. Halmshaw: I would like to suggest that the data given for fluoroscopy in Tables 2A and 2B have been obtained under somewhat abnormal conditions. Using an ultra-fine-focus fluoroscopic tube, with suitable projective magnification, and a screen-tube distance of 20 in, D.I.N. wire-penetrameter sensitivities of about $1\frac{1}{2}\%$ can be obtained on aluminium thicknesses between 1 in and $2\frac{1}{2}$ in, and 3% wire-penetrameter sensitivity is possible through 1 in of steel. These figures are not given with any intention of decrying the image intensifier; they are only possible by working in complete darkness and after very thorough dark-adaptation; such conditions would not be possible in many industrial applications.

In using the image intensifier I have noticed that the 'boiling' of the image, which I am told is due to quantum fluctuations, commences at quite high brightnesses, and I wonder whether this effect is more likely to prove the major factor limiting the sensitivity attainable with the apparatus than screen blurring or scattered radiation in the specimen. On the other hand, calculations of Coltman and others suggest that the theoretical limit of sensitivity due to quantum fluctuations is considerably higher than the sensitivities at present attainable.

For the fluoroscopy of small specimens I have found that the radiation reaching the screen round the specimen must be eliminated or reduced; a lead mask cut to fit the specimen, as used in Figs. 5 and 6, is efficient but cumbersome, and I have recently found that a foil of 0·006 in lead placed between the specimen and screen—not a mask, but a filter across the whole screen—results in a very considerable improvement in image detail. One advantage of using an ultra-fine-focus X-ray tube with the image intensifier is that it is possible to alter the amount of projective magnification so that the image of the specimen fills the screen; the degree of magnification is not very critical, as is shown in Fig. 4.

Mr. W. Bamford: The image-intensifier tube dealt with in the paper is of considerable interest, and I feel that justice has not been done to its technical description. It is quite a paradoxical device in the sense that the actual image is first of all reduced in order that it may afterwards be magnified. Fig. 1 shows a very simple construction, and I should like it to be confirmed that there are no additional electronic focusing plates or rings between the photo-emissive screen and the viewing screen. Additional information about the optics would be appreciated.

Whilst there are other methods of dealing with X-ray diffraction patterns, it occurred to me that some adaptation of this image intensifier tube might be used for magnifying such images. They are inevitably very small and of low intensity, and consequently the method adopted in the image-intensifier tube could not be directly applied. Is it possible that suitable tubes could be put in series so that, instead of viewing the first intensified image, this could be allowed to fall on a second photo-emissive screen so as to produce a further degree of amplification?

Mr. J. J. Bliss: There are many instances where fluoroscopy

is not accepted for examination of components, since existing regulations demand that a permanent record of each unit examined shall be retained. Examination of shell fuses is required to check that the internal components are correctly assembled and not to detect small flaws in the material; thus fine detail is not important. To reduce costs, sensitized paper is commonly used in place of film, since it gives adequate definition for the purpose. If micro-film fluorography with the image intensifier were introduced, a saving in film, processing and storage costs might be achieved. Against this saving, however, one must set the increased number of exposures and the modification to the handling gear involved by the change from a large sheet of paper, accommodating many fuses, to the 5 in diameter circle of the image intensifier accommodating, perhaps, only one.

The second case involves the production of some components, such as turbine rotor blades, where rejections may be as high as 50% and where a full-size radiograph of each part passed into service must be retained. Here it has been found that fluoroscopy with the image intensifier gives sufficient detail for some 30% of the blades to be rejected before radiography. Thus a saving of some 30% in film and processing costs is obtained, while no blade is passed without a full-size radiograph being taken of it.

Finally, I would like to ask the authors whether they have yet used ciné-fluorography in connection with industrial work. Last week the British Institute of Radiology devoted a morning of their Congress to papers and films on ciné-fluorography with the image intensifier. Some remarkable films were shown. It would appear that movements, even when a certain amount of definition is sacrificed, can bring out details that might otherwise have passed unnoticed in an ordinary radiograph, which is a static and momentary view of a dynamic process. There may be similar examples in industrial work.

Mr. G. Syke: Could the authors please indicate the approximate dose-rate required at the window or fluorescent screen of the image intensifier in order to produce a conveniently viewable image? I ask this question with some thoughts about the possibility of using the image intensifier for fluoroscopy with radioactive sources.

If it is effective with the gamma radiation of sources like caesium 137, iridium 192 or cobalt 60, it might considerably extend the scope and usefulness of radiography and even more of fluoroscopy with radioactive sources in the heavier thickness ranges for which X-ray sets are very costly.

An earlier speaker has referred to the inspection of shell fuses and to the economics of this operation. A few years ago I was associated with the development of a gamma-ray apparatus for this very purpose. The task was to ascertain whether a critical component inside the assembly was present and in its correct position. Each fuse was placed in a fixture and the attenuation of a narrow beam of gamma rays along a critical path through the specimen was measured. Two such discreet measurements gave the full answer. The results of the measurements are easy to interpret and lend themselves very conveniently to being put down on a strip chart recorder for subsequent evidence. This technique can be readily extended to fully automatic inspection where a large number of standard specimens are involved.

Mr. S. N. Pocock: The industrial user is naturally interested in the economies which can be effected by the use of this new instrument. In considering these, however, he is obviously concerned with certain aspects of the actual tube and its maintenance, and I should like to have some clue as to the life of the tube and the approximate cost per hour resulting from that limited life. Also, can appreciable differences in definition be expected due to variations between one tube and another, and how does the tube approach the end of its life? Is this clearly visible in any way, or is the falling-off in performance so gradual as to represent a hazard to the efficiency of inspection?

THE AUTHORS' REPLY TO THE ABOVE DISCUSSION

Dr. A. Nemet and **Mr. W. F. Cox** (*in reply*): Several speakers suggested that the intensifier tube should have been described in more detail. Since the length of the paper is limited we have reviewed the tube briefly in order to give more space to original results. The tube has been described in Reference 1 and also at a discussion before The Institution.*

There were also several questions about statistical fluctuations. The absorption in the X-ray screen of 40 kV X-radiation filtered with Bakelite phantom is about 25%. The design, we think, represents a good compromise between statistical fluctuations, intensification factor and blurring up to 125 kVp.

In reply to Dr. Taylor, equipment is available for rapid change-over between fluoroscopy and radiography and is described by P. M. van Alphen.†

With regard to fluoroscopy of spot welds, the usefulness of the intensifier will depend on the defect that can be tolerated. Hair cracks, however, are extremely difficult to detect even with radiography.

Referring to grain size, resolution is not always limited by the film even in the case of miniature radiography. This agrees with Mr. van Someren's observations. In reply to Mr. van Someren we think that, in certain circumstances, cracks in even large objects may more easily be detected by fluoroscopy than by radiography due to movement.

In reply to Mr. Rockley, all our practical results for aluminium were obtained with specimens under 1·5 in, but our sensitivity measurements extend to 7 in.

In reply to Mr. Allwood, we have found that the resolution of the intensifier itself and also the sensitivity measurements are uniform across the image field except for a distance of 0·25 in from the edge.

Mr. Halmshaw's remarks on the use of the lead filter to improve detail are very interesting and may offer a useful practical means of improving image quality and avoiding masking, particularly when used in conjunction with a diaphragm.

In reply to Mr. Bamford, the tube has been tried for diffraction work, but its value in this technique is limited by the absorption in the window. In this connection we should like to mention that the well-known unwanted *Laue* patterns often found during radiography of stainless steel are seen quite strikingly on image-intensifier fluoroscopy. We have found, however, that when using ×2 magnification it is very easy to distinguish between the diffraction spots and the flaw by slight rotation of the object. The spots will travel much faster than the flaw and will disappear.

We were interested to hear the remark of Mr. Bliss on the advantage of the intensifier tube in the examination of rotor blades since, indeed, this is an application where the special properties of the new technique lead to direct economy compared with radiography. We have not so far had applications of industrial ciné-radiography but we expect that there are a number of problems where both movement examination and permanent record are required.

In reply to Mr. Syke, we would estimate the dose rate at the fluorescent screen to be approximately 10–20 mr/min, for approximately 50 kV radiation. With regard to the sensitivity of the tube for γ-radiation from isotopes we do not think that

* See *Proceedings I.E.E.*, 1955, 102 B, p. 845.
† 'Optical Aids for the Image Intensifier', *Philips Technical Review*, 1955, 17, p. 77.

he present tube design is suitable to produce useful fluoroscopic esults at gamma energies of the megavoltage region. For hulium we estimate about 20 curies are necessary for fluoroscopy of 1 in aluminium at a distance of 15 in.

In reply to Mr. Pocock, modern tubes should give a useful ife of at least a year. A feature of the intensifier is that when he tube approaches the end of its life the definition does not leteriorate; instead, a bright spot appears in the centre, which grows in size and intensity with age. There is thus no hazard o the efficiency of the inspection due to a falling-off in per- ormance, and a timely warning is given.

With regard to variations between individual intensifier tubes,).3 mm blurring is the maximum manufacturing tolerance.

Lastly, we should like to deal with the important question of nandling, to which several speakers have referred. It should be appreciated that all our measurements were carried out with a simple research type of equipment such as was demonstrated in the lecture theatre. We think that a unit designed on this principle is flexible enough for a development laboratory. We are aware that, for practical problems in industry, handling gear designed in connection with the intensifier is of paramount importance, and that the value of the method will depend to a large extent on the quality of the mechanical engineering applied to the problem. The example taken from the German steel-tube factories shows how such a problem can be satisfactorily solved (Fig. 8) for a specific production problem. Clearly the number of similar examinations will decide the extent of tooling required.

We are very glad that the discussion has brought out some valuable suggestions and comments which we feel may also be useful to the designer of handling equipment. We think that, as this technique matures, the image intensifier may become a valuable addition to the methods of non-destructive testing.

621.317.7 : 621.398 : 621–5 : 621.771

Paper No. 1933 M
Nov. 1955

THE INSTRUMENTATION OF A 14-INCH EXPERIMENTAL ROLLING MILL

By S. S. CARLISLE, M.Sc., Associate Member, and G. W. ALDERTON, B.Sc.

(*The paper was first received 18th January, and in revised form 6th August*, 1955. *It was published in November*, 1955, *and was read before a joint meeting of the* MEASUREMENT AND CONTROL SECTION *and the* UTILIZATION SECTION *6th December*, 1955, *and the* NORTH-WESTERN MEASUREMENTS GROUP *24th January*, 1956.)

SUMMARY

The paper describes the comprehensive instrumentation scheme recently developed and applied on the new 14 in experimental rolling mill in the British Iron and Steel Research Association (B.I.S.R.A.) Sheffield Laboratories. Roll force on each side of the mill, front and back tensions and strip gauge, using the B.I.S.R.A. gaugemeter principle, are continuously indicated on high-speed servo-operated potentiometric indicators fitted with large pointers and dials. The supply voltage to all measuring elements and transducers in the system is at a frequency of 400 c/s and a rational system of electronic units has been worked out which enables identical amplifiers to be used for all channels. Some of these amplifiers are already in quantity production for other purposes.

In the a.c. gaugemeter system used the normal gauge-setting control is situated on the main control desk, and a device is included for automatically setting the zero of the gauge-deviation indicator. Also the mill-screw setting data are fed into the gaugemeter system by Magslip transmission from the screw top. The five indicating dials are mounted in one unit on the mill housing. The whole instrumentation scheme is designed and laid out as a model system for industrial mills of this class.

Facilities are provided for experiments with various systems of automatic gauge control previously developed by the B.I.S.R.A.[1] A simple method of on/off control of the existing a.c. screw motors according to the gauge deviation has been successfully operated with a tolerance zone of only $\pm 0{\cdot}0005$ in.

Facilities have also been provided for switching to automatic tension control based on measured tensions as an alternative to the present method of control of coiler-motor current with correction for radius of build-up.

The accuracy of the various measuring systems is discussed in the light of experience gained so far, and attention is drawn to possible sources of error, particularly in the gaugemeter system.

(1) INTRODUCTION

This mill has been installed to facilitate research into cold rolling at high speeds and in particular for experiments with new systems of automatic gauge control. It is briefly described as a 14 in four-high reversible mill equipped with coilers by which front and back tension can be applied. The essential features of a mill of this type are shown in Fig. 1. The system of four rolls (four high) is used, so that small-diameter rolls can be employed to roll thin strip efficiently, and the stiffness necessary to prevent roll bending is provided by the large-diameter back-up rolls. The force applied to the rolls, and hence the reduction in gauge effected by a pass through the mill, is controlled by the motor-driven screws which apply the downward force on the roll bearings. The coilers and uncoilers on each side of the mill are provided with adequate power so that appreciable tension can be set up in the strip on both sides of the mill. This is done because both front and back strip tension affect the rolling process. Since it is necessary that the maximum tension allowed —5 tons in this case—be provided right up to maximum mill speed, it is clear that the coiler motors must have a very high power rating. The mill is designed to operate at strip speeds of up to 2000 ft/min. To provide constant tension during an experiment it is necessary that the coiler torques be automatically corrected for build-up of strip on the coilers, which alters the effective radius. The control equipment and the mill drives are arranged so that it can roll strip in either direction. This feature greatly facilitates the use of the mill for experimental purposes and is representative of a large number of mills operating in industry.

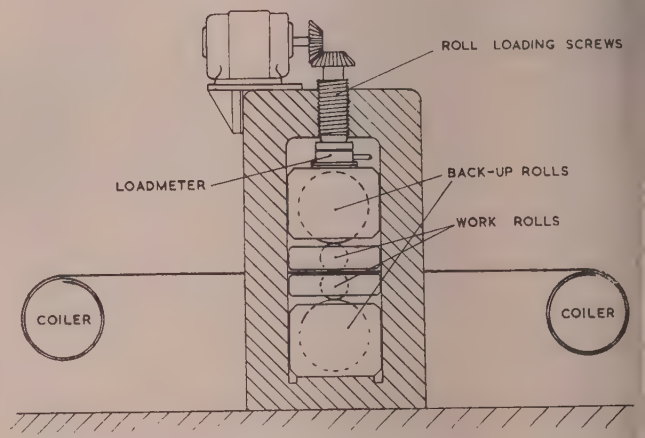


Fig. 1.—Arrangement of the 14 in four-high reversible cold rolling mill.

The essential variables in the rolling process are roll force (applied on each side of the mill by the screws), strip tension both front and back, strip gauge on leaving the mill and the driving torque at the mill rolls. It is necessary that the majority of these variables should be indicated and recorded to facilitate operation of the mill and research into rolling processes. Methods had been developed earlier by the British Iron and Steel Research Association (B.I.S.R.A.) for measurement of roll force,[2] strip tension[3] and strip gauge.[4] These methods all made use of electric resistance strain gauges for the primary measurement, but a variety of circuit techniques were employed according to the application. It was decided to use these principles in the instrumentation of this mill, but to unify the design of the instrumentation system and introduce circuit techniques and methods of display which would act as a model upon which future mill instrumentation schemes in the industry might be based. Special attention was given to rationalization of the design of the various measuring systems to facilitate their commercial reproduction and to make easier the proper servicing of the equipment.

The electronic apparatus employed with the measuring systems also provides the basis of a variety of systems of automatic gauge control. One system of control has already been successfully operated, using the gauge indicating pointer to provide on/off control of the mill screw motors via their existing contactors.

The paper is based on Report No. MW/A/32/54 of the Rolling Committee of the British Iron and Steel Research Association.

Mr. Carlisle is at the South Wales Laboratories of the British Iron and Steel Research Association.

Mr. Alderton is at the British Iron and Steel Research Association Metal Working Laboratories.

(2) BASIC MEASURING SYSTEM

During recent years a satisfactory system of measurement of roll force in rolling mills has been evolved by B.I.S.R.A.[2] using steel load cylinders fitted with electric resistance strain gauges and located under the mill screws. The B.I.S.R.A. gaugemeter[1] also employs these load cylinders. The method of measurement of strip tension, either by torque measurement in the coiler shafts with suitable correction for coil radius build-up,[3] or by a three-roller system over which the strip is passed,[5] is also basically dependent on electrical resistance-strain-gauge bridges.

Since a strain-gauge bridge forms the source of the measuring signal in all three measurements, it was decided to evolve a circuit technique and system of amplifiers which would be generally applicable. The basic circuit technique chosen is shown in Fig. 2. It employs a well-known automatic potentiometric

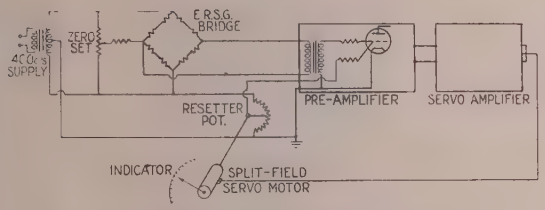


Fig. 2.—Schematic of potentiometer indicator system.

system of indication in which the output signal from the bridge is compared with a signal generated by a linear potentiometer, and any difference signal is amplified and used to drive a servo motor coupled to this resetting potentiometer. The system thus operates until a condition of balance is achieved in which the angular position of the resetting potentiometer is a continuous measure of the magnitude of the unbalance of the resistance-strain-gauge bridge. The following features of the system are noteworthy:

(*a*) The use of a self-balancing potentiometric system of indication makes the calibration independent of the gain of the electronic amplifiers employed. This contributes to reliable operation, easier installation and initial adjustment, and simplifies servicing. Furthermore, a mechanical drive of appreciable power is available to operate a large and easily read pointer. This feature is particularly advantageous for industrial instrument displays.

(*b*) The bridge is excited from a 400 c/s source. This a.c. excitation permits the use of straightforward electronic amplifiers and avoids the many problems introduced in the amplification of low-level d.c. signals. Also any errors due to thermal e.m.f.'s such as might arise with a d.c. circuit technique working at very low signal levels are avoided. The frequency of 400 c/s has been chosen as a carrier in preference to 50 c/s in order to provide the quick-response system which will ultimately be required in certain systems of automatic gauge control. This higher frequency has also the advantage that discrimination against 50 c/s pick-up is easily provided. It is also an accepted standard for a very wide range of servo mechanisms, and already a large number of servo amplifiers designed to work on this frequency are available and can be directly employed in this measurement system.

(*c*) The centre point of the bridge supply transformer is earthed, and, by placing an earthed screen round the bridge-output connecting wires, leakage currents across the bridge arms, which might cause unbalance, can be prevented. This technique of guarding is particularly desirable for high-resistance bridge circuits, but is only an added precaution for the low-resistance foil gauges actually used in this system.

(2.1) Roll-Force and Tension Measurement

The circuit shown in Fig. 2 includes the essentials of the roll-force measuring system. For tension measurement, the method using coiler-shaft torque suitably corrected for coil radius has been employed. The basic circuit is shown in Fig. 3. The voltage from the resetting potentiometer is attenuated by the

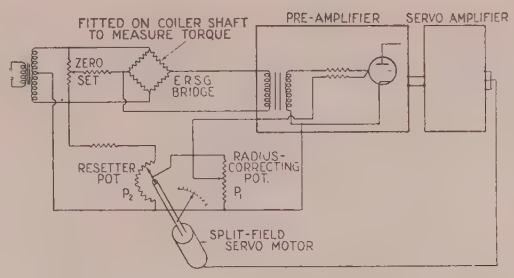


Fig. 3.—Basic circuit of strip tension measuring systems.

potentiometer P_1 according to the radius of the coil. It can be shown that the pointer indication is thus proportional to the output signal from the resistance strain-gauge torque bridge divided by the coil radius. Thus the indication is proportional to strip tension.

(2.2) Strip-Gauge Measurement

For continuous measurement of strip gauge the B.I.S.R.A. gaugemeter principle[1] has been employed. The operating principle of this instrument is that the introduction of a strip into the roll gap causes an elastic deformation of the mill housing, rolls and screws, and thus a continuous measurement of roll force will give a continuous indication of variations in strip gauge, provided that the screw setting is fixed. Hence by combining the signal from a roll-force meter with a signal representing screw setting, an absolute measure of strip gauge can be made, unaffected by variations in the screw setting. Fig. 4

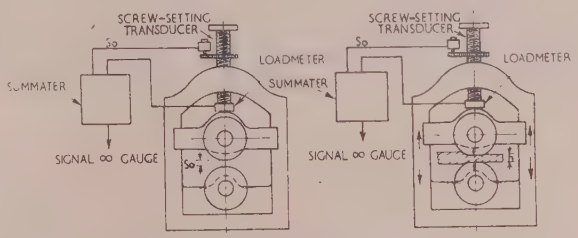


Fig. 4.—Principle of the gaugemeter.

S_0 = Gap between volts with no strip present.
h = Exit gauge = Original gap S_0 plus amount of stretch of mill housing and compression of rolls.

shows the principle. Suppose that the setting of the screw is such that there is a gap S_0 between the rolls with the mill unloaded. When the mill is loaded by strip between the rolls, a roll force F is generated, and if the coefficient of elasticity of the mill housing and rolls is M, the gap between the rolls will increase by F/M owing to the presence of the strip. If the strip exit gauge is h

$$h = S_0 + \frac{F}{M}$$

Thus it is only necessary to add together a signal proportional to S_0, which can be generated by a suitable transducer operated from the mill screws, and a signal from the roll-force meter, which is proportional to F. The constants of the screw-setting transducer and the roll-force loadmeters must be matched so that their output signals are in the correct ratio for the mill elastic coefficient M.

The basic circuit technique used is shown in Fig. 5. Again a.c. excitation has been used in order to unify the electronic equipment required.

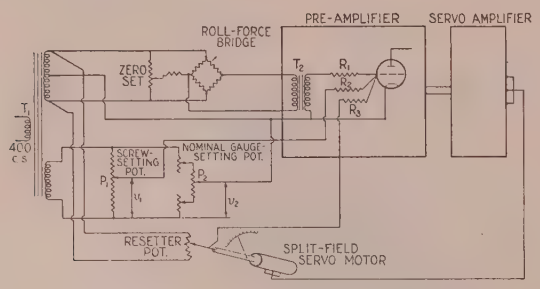


Fig. 5.—Basic circuit of a.c. gaugemeter system.

A signal proportional to roll force is provided in the usual manner and fed to the pre-amplifier via a step-up transformer, as in the circuits described earlier.

The gaugemeter is designed so that the indicator shows only the deviation in gauge about a nominal value which can be set on control knobs. This is done by using a double-decade potentiometer P_2 connected in parallel with potentiometer P_1, the screw transducer. The mechanical drive from the mill screw to potentiometer P_1 can be matched to the resistance steps of the double-decade potentiometer P_2 so that each step of P_2 corresponds to a certain change in screw setting measured in thousandths of an inch.

A signal corresponding to the deviation in gauge about the set nominal value is obtained by adding the voltage difference $(V_1 - V_2)$ to the roll-force signal determining the stretch of the mill housing. This summation is done in the pre-amplifier by the resistors R_1 and R_2. As in the previous circuits this resultant signal is measured by balancing it against a signal from a resetting potentiometer added in through summating resistor R_3. When the resetting potentiometer has been driven to a position of balance, its indication represents gauge deviation about the nominal value set on the double-decade potentiometer P_2.

In the design of this circuit particular attention has to be given to the transformer T_1, so that its secondary voltages are maintained accurately in phase with each other for the particular loads applied to the windings. Also the transformer T_2 and circuit of the resetting potentiometer have to be designed to minimize phase shift, so that the three voltages summated in R_1, R_2 and R_3 are held accurately in phase.

The circuit shown in Fig. 5 is much simplified and does not show adjusting and calibrating controls. These will be discussed later in relation to the actual circuit diagrams of the various sections of the equipment.

(2.3) Amplifiers and Control Units

The circuits of the pre-amplifiers and servo amplifiers employed in the systems described for roll force, strip tension and gauge measurement are almost identical, their only individuality being in the voltage gain required to accommodate the different levels of input signal and in the gaugemeter pre-amplifier, where three summating resistors are used instead of two.

Beside the units so far mentioned, another termed a "control unit" is introduced, whose function is to accommodate the zero-setting and calibration controls and the transformer for supplying the appropriate resistance-strain-gauge bridge in each measuring system. The control unit also serves as a junction unit to permit a good wiring layout.

The general arrangement of the instrumentation scheme is shown in Fig. 6. The arrangement of the roll-force and tension measuring systems will be clear from previous discussion. The control units for these measurements are built as two-channel units. The gaugemeter control unit, in addition to housing the necessary zero-setting and calibration controls, also includes the screw-setting potentiometer with its servo-motor drive which is operated from a Magslip transmitted on the mill housing. This remote control of the screw potentiometer by a Magslip system simplifies the fittings on the mill housing and locates the potentiometer in a more accessible position for servicing.

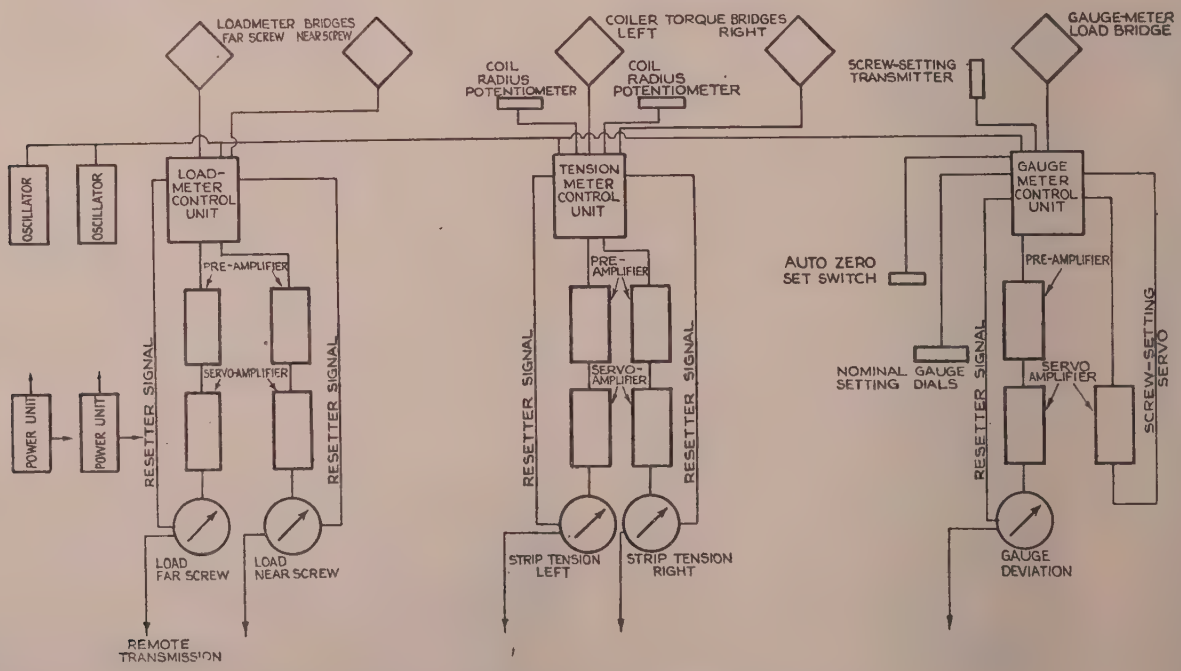


Fig. 6.—General arrangement of the 5-channel instrumentation system.

The gaugemeter control unit also includes an automatic zero-setting mechanism, the operation of which is facilitated by the use of a servo drive to the screw potentiometer. Fig. 6 shows how the whole instrumentation system lends itself to standardization of the electronic apparatus employed.

(3) CIRCUIT DETAILS OF THE EQUIPMENT

Fig. 6 shows the essential units of the system. There are five pre-amplifiers, six servo amplifiers and three control units, two oscillator units to supply the 400 c/s power supply and two power units to provide the necessary d.c. supply for the whole equipment. In addition, a switching unit has been included to centralize the switching control of the apparatus.

All the electronic equipment is contained in a multi-unit rack, which was designed to house servo amplifiers in a complex system of electronic servo mechanisms designed for the Admiralty; it has been found very useful in this application to a multi-channel mill instrumentation scheme. The individual electronic units fit into this rack and make contact via plugs and sockets. The servo-amplifier units are built on trays, 3 in wide by 8 in high and 17 in long, and are of a type now being produced in large quantities. The oscillators, pre-amplifiers and load- and tension-meter control units are built on to trays identical with those on which the servo amplifiers are assembled. The power units and gaugemeter control unit are built on heavier sliding trays which occupy the space of three amplifier trays.

(3.1) The Pre-Amplifiers

Fig. 7 shows the circuit of the pre-amplifier unit, which provides a voltage amplification of up to 500 and whose output is designed to feed into the servo amplifier. The pre-amplifier input circuit includes the three summating resistors with accommodation for a fourth. The amplifier includes an 8 : 1 step-up transformer suitable for receiving a signal from a resistance-strain-gauge bridge (with 100-ohm output impedance) source. Between the second and third stages of the amplifier a bridged-T filter network is introduced which is tuned to reject 50 c/s signals. This has been found useful in eliminating troubles due to 50 c/s pick-up. The filter network, however, introduces a phase shift at 400 c/s, and correction has to be applied in other parts of the amplifier in order to maintain a relatively small phase shift between amplifier input and output at 400 c/s. The performance of the amplifier is thus frequency-sensitive, but not critically so, and a variation of $\pm 5\%$ in frequency does not cause any appreciable change in performance. The supply to the amplifier is 400 volts d.c. and 6·3 volts for valve heaters.

The residual signal at the amplifier output with all the inputs short-circuited is less than 10 μV referred to the input. The gain of the amplifier is adjusted to suit the particular application. In the roll-force measuring circuits the amplifier is only required to give a gain of approximately 50, while for the gaugemeter circuit a gain of 150 is used.

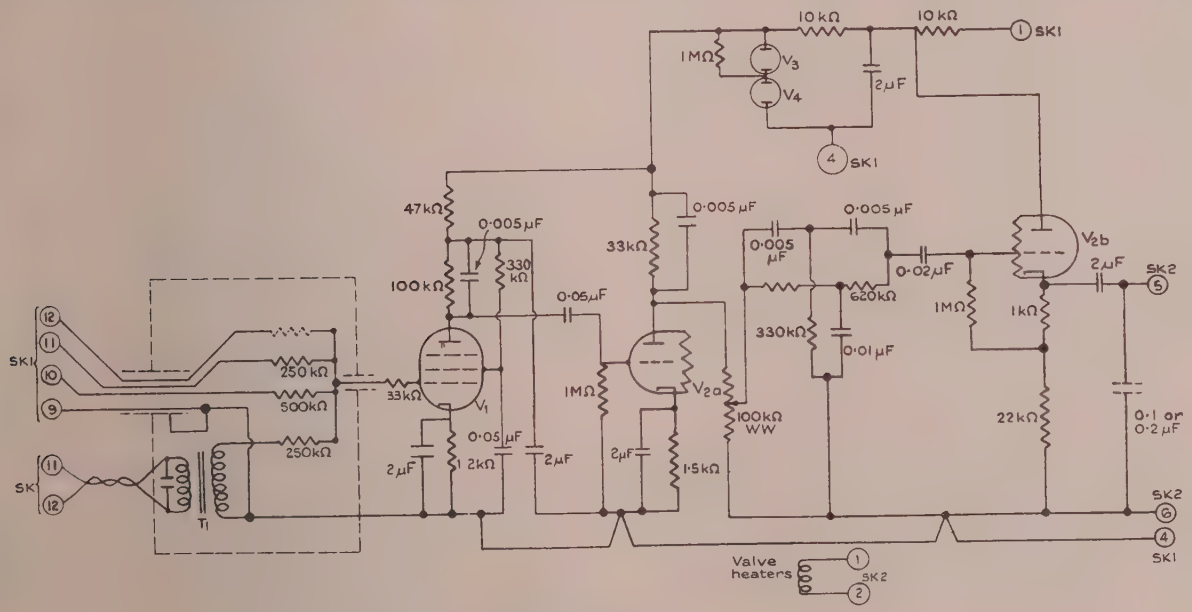


Fig. 7.—Pre-amplifier unit.

(3.2) Servo-Amplifier Unit

The servo amplifier was developed by the Admiralty and is now being manufactured in quantity by several firms. The only modification necessary to use it in this system is a change-over from 400 c/s voltage supply to the valve heaters to a 50 c/s supply.

The function of the amplifier is to receive a 400 c/s signal, rectify it in a phase-sensitive rectifier and generate a d.c. output of sufficient power to drive a split-field servo motor. The amplifier includes facilities for the use of phase advance for stabilization of the servo mechanism in which it is employed and also for injection of d.c. feedback signals as an alternative means of stabilization. The amplifier can be adjusted to give such a gain that a 10 mV 400 c/s signal at the input will give a maximum differential current output of 80 mA. This gain is available only if the phase-advance circuit is not used. With phase-advance stabilization the gain is such that 100 mV is required at the input to produce maximum output.

(3.3) Oscillator Unit

In order to supply the necessary power at 400 c/s to excite the resistance-strain-gauge bridges, the mill-screw Magslip trans-

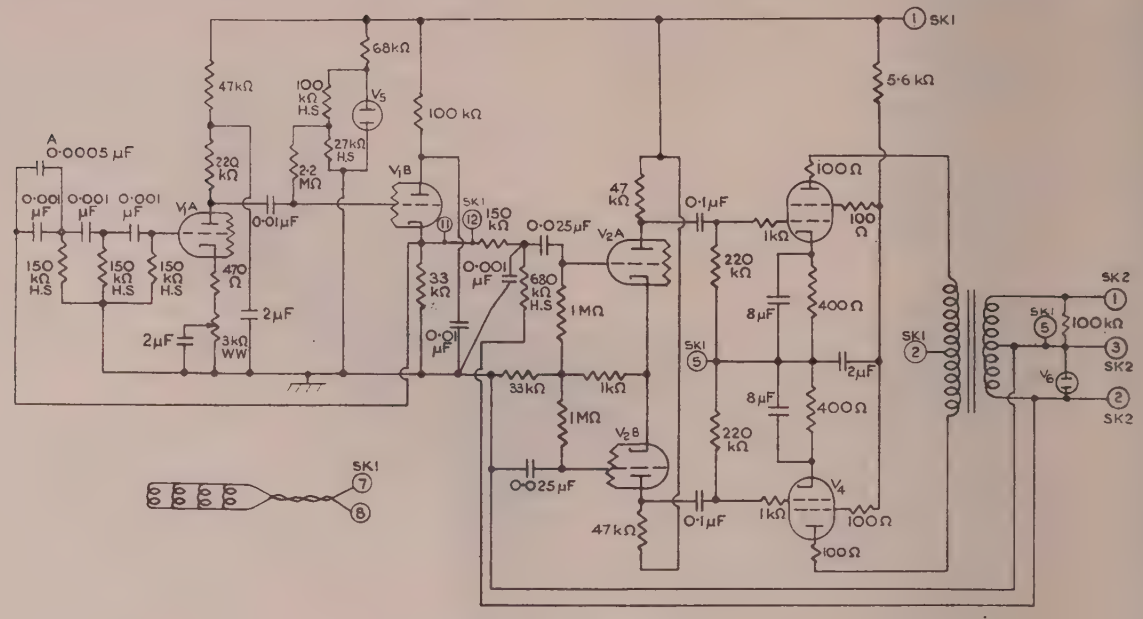


Fig. 8.—Oscillator unit.

Zero-set controls are on front panel.
SK1—Far loadmeter. SK2—Near loadmeter.

mitters, the components in the gaugemeter circuit and the reference frequency to the servo amplifiers, a suitable oscillator was developed. The circuit is shown in Fig. 8. It is of orthodox design using a resistance-capacitance phase-shift circuit to define the frequency. The amplitude of the oscillator output is stabilized by limiting at the bottom bend of the cathode-follower characteristic. Feedback is provided from the secondary of the output transformer into the input of the amplifier stages in order to improve the regulation of the oscillator and reduce harmonic distortion. The unit is designed to give an output of up to 20 watts at 115 volts 400 c/s.

The use of the potentiometric system already described makes the calibration of the equipment independent of appreciable variations in amplifier gain and 400 c/s supply voltage. It is, however, desirable that, should the oscillator voltage drop by, say, 30%, it will be indicated so that the equipment can be inspected. A small neon glow-lamp connected across the oscillator output indicates normal functioning of the unit.

(3.4) Power Supply Unit

The power supply unit provides an output of 550 mA at 400 volts d.c. employing full-wave metal rectifiers. A well-smoothed lower-current supply is provided for the pre-amplifiers. Separate low-current 6·3-volt heater supplies are provided for the rectifier stages in the servo amplifiers.

The unit is built on a servo tray which slides into the rack and occupies the space of three amplifier chassis.

(3.5) Loadmeter Control Unit

The circuit of the loadmeter control unit is shown in Fig. 9. It acts as a junction unit for connections to the two roll-force load cylinders, to the pre-amplifier and to the potentiometric indicator unit. It contains a transformer fed with 115 volts at 400 c/s from the oscillator, and supplies the requisite voltage to the two electrical resistance-strain-gauge bridges measuring roll force. Also included are the resistance and capacitance zero-setting controls for the two bridges and the controls for adjustment of the calibration of the two measuring channels.

Foil resistance strain gauges are used on the load cylinders. These gauge elements are of 50 ohms resistance each and are used in an 8-element bridge with two elements in each arm, which form a 100-ohm load.

The resetter potentiometer is fed through an adjustable potential-divider chain to provide means of setting the calibration. The bridge output voltage for 100 tons load on each cylinder is approximately 0·7 mV per volt. This defines the resistances in the potential-divider chain feeding the resetter potentiometer. This potentiometer is shunted with a 100-ohm resistor in order to maintain as low an output impedance from the potentiometer wiper as possible. Particular attention has been given to the maintenance of low circuit impedances in the signal circuits of all parts of this instrumentation system where capacitive loading by long connecting cables might cause phase shift of the signals. All the fixed resistances employed in the system are provided by high-stability resistors.

The control unit is built on a plug-in servo-amplifier chassis.

(3.6) The Tension-Meter Control Unit

The general form and function of this unit is similar to the loadmeter control unit. The basic circuit of the tension-measuring system has already been shown in Fig. 3. Fig. 10 shows the circuit detail of the two-channel control unit associated with this part of the system. The particular feature of this control unit, as distinct from the loadmeter control unit, is that the resetting-potentiometer output is attenuated according to the setting of the coil-radius-correcting potentiometer before being fed into the pre-amplifier. This provides for the necessary division of the torque signal by a factor proportional to the coil radius to give the strip tension. Also the resistance network feeding the resetter potentiometer is arranged so that the zero-potential point on this potentiometer is a short distance from the end of the winding. This is necessary to facilitate setting up and calibration of the tension-measuring system.

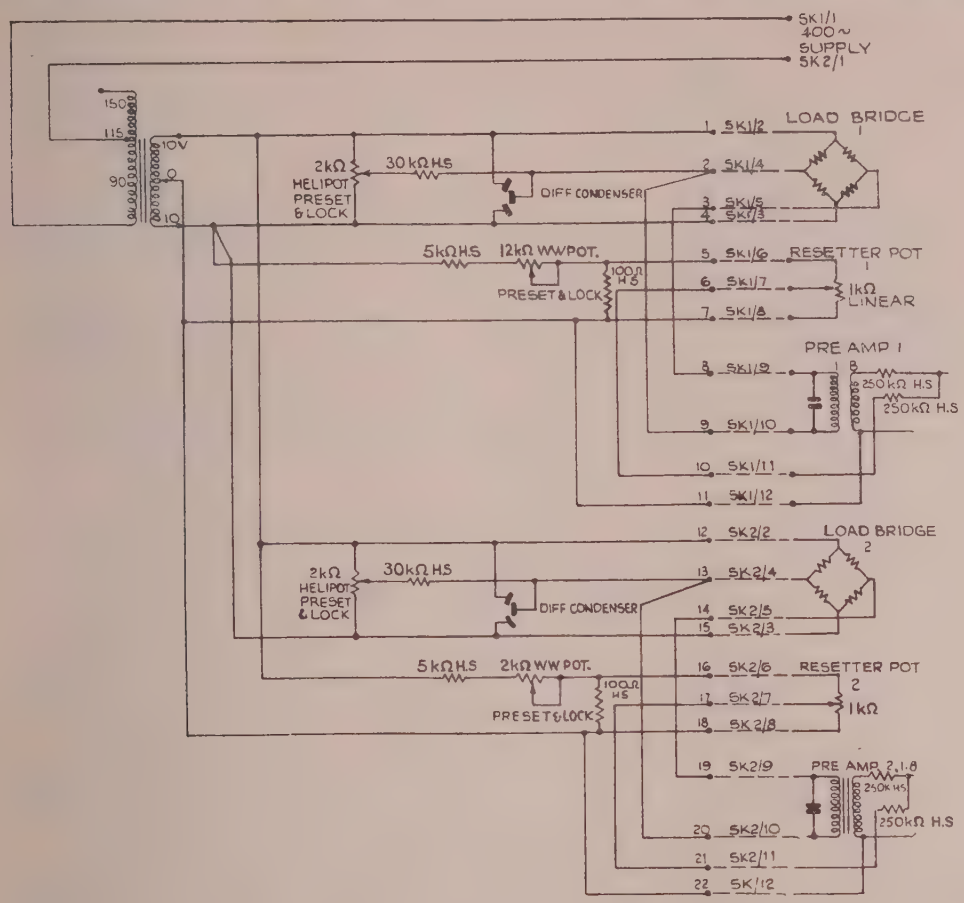


Fig. 9.—Loadmeters control unit.

Resistance-strain-gauge elements of 5 000 ohms resistance each were already fitted on the coiler shafts for torque measurement before this instrumentation scheme was installed, and it was therefore decided to use these high-resistance bridges temporarily until low resistance elements could be fitted. The high-resistance system necessitates a different bridge-supply transformer from that used in the loadmeter control unit and a different input transformer ratio in the pre-amplifier (2 : 1 step-down into the amplifier).

The radius-correcting potentiometer is a specially made linear potentiometer of 15 kilohms total resistance. Its wiper is moved along the winding by a push rod fitted with a roller which bears on the surface of the coil, so setting the potentiometer according to the coil radius. The full traverse of the potentiometer corresponds to a coil-radius change of from 7 to 28 in. The 400-ohm fixed resistance and the 100-ohm variable resistance in series with the radius-correcting potentiometer are introduced and adjusted so that a given radius change causes exactly the right percentage change of reading of tension. The actual calibration of the tension meter in terms of movement of pointer per ton of tension is carried out by the 30-kilohm variable resistance in series with the radius-correcting potentiometer.

The procedure for setting up this measuring system is first to adjust the bridge zero-set potentiometer until the bridge is accurately balanced with zero torque. This will cause the potentiometric indicator unit to move its resetting potentiometer to the true electrical zero and the pointer zero is made to coincide with this. A fixed torque is now applied to the coiler shafts, giving a tension reading. The ratio of the tension indicated for two measured settings of the radius correcting potentiometer is compared with the ratio of the corresponding radii. By adjustment of the 100-ohm series resistor these two ratios can be made to agree. The calibration of the tension meter can then be adjusted accurately by applying a known strip tension and adjusting the 30-kilohm series resistor until the correct reading is given.

A $0{\cdot}2\,\mu$F condenser is connected across one-half of the supply-transformer secondary winding. This capacitive load is introduced in order to match accurately the phases of the voltages on the two halves of the secondary winding. Without the capacitance correction they are slightly out of phase, owing to differences in the winding characteristics of the two halves.

All the fixed resistors used in this unit are high-stability components.

(3.7) The Gaugemeter Control Unit

The circuit detail of this unit follows from the basic circuit shown in Fig. 5. The top half of Fig. 11 shows the essential gaugemeter circuit. The lower half, which will be described later, is concerned with a servo-mechanism drive for the screw-setting potentiometer.

The function of this circuit is to combine in a suitable ratio a signal corresponding to roll force, which is a measure of the mill housing stretch, and a signal corresponding to the setting

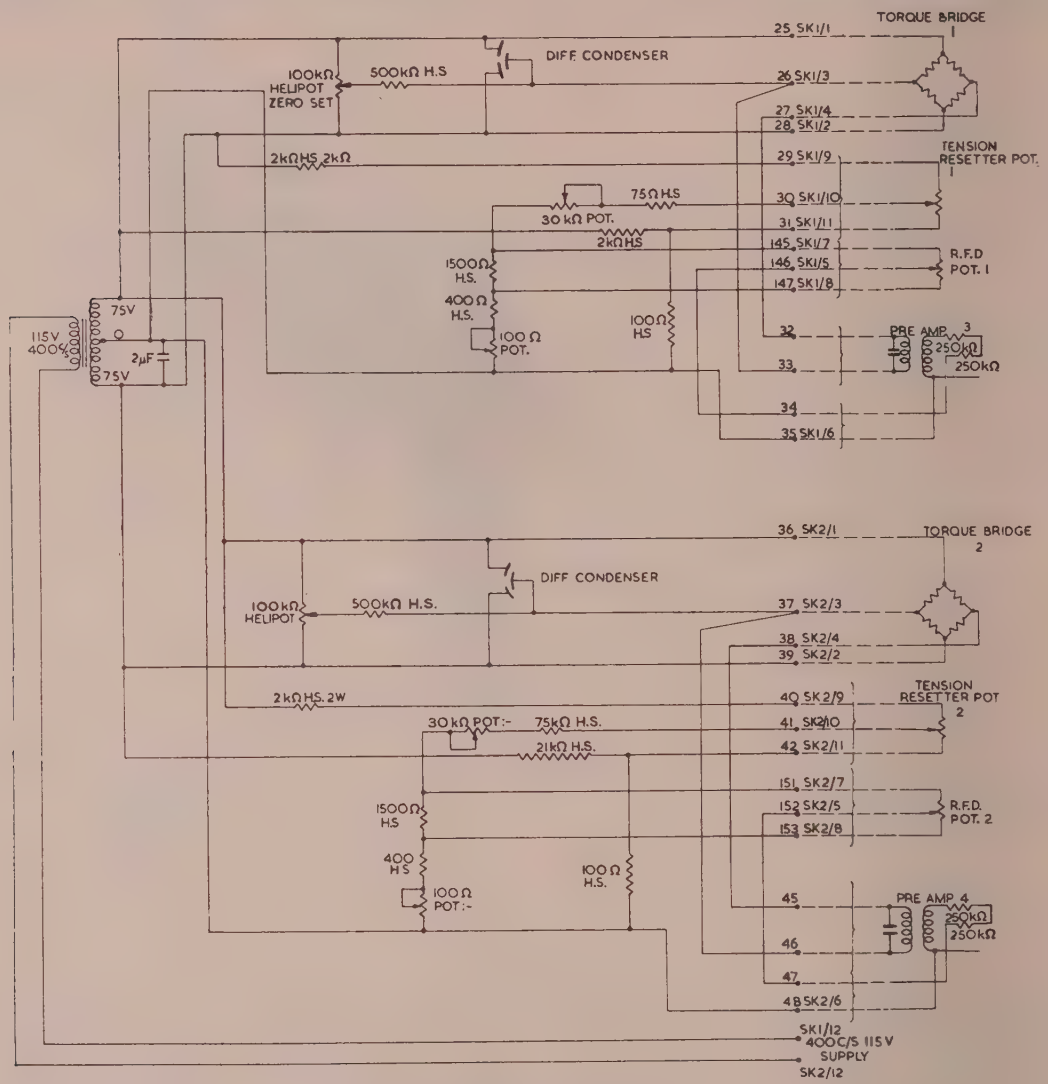


Fig. 10.—Tension-meters control unit.

SK1—Left tension meter.
SK2—Right tension meter.

of the mill screws. The combination of these signals gives a measure of the strip gauge in the roll gap. The resultant signal is displayed as a deviation about a nominal gauge set on a double-decade potentiometer (gauge-setting unit).

The calibration of the gaugemeter circuit in terms of millivolts per mil of gauge is basically determined by the sensitivity of the roll-force measuring load cylinder and the excitation voltage applied to its resistance-strain-gauge bridge. The load cylinders used, which were fitted with 50-ohm foil gauges, give an output of approximately 0·007 mV per volt excitation of the bridge per ton load applied to the cylinder. The coefficient of elasticity of the mill housing and rolls was measured and found to be approximately 3 000 tons/in. Thus a load of one ton on each side of the mill will produce a stretch of 2/3 000 in, and this stretch will correspond to 0·007 mV per volt applied to the load-cell bridges. Assuming a bridge excitation of 20 volts, the calibration of the system is approximately 0·21 mV/mil.

Since the bridge transformer in the pre-amplifier has a ratio of 8 : 1 the signal at the summating resistors is 1·7 mV/mil.

The screw-setting potentiometer is coupled through a Magslip servo mechanism to the mill screw through a gear ratio such that the winding length of the screw-setting potentiometer represents 150 mils of screw movement. This was considered sufficiently large because the mill will not be required to roll material having a thickness greater than about 0·1 in. Thus if this potentiometer is to represent 150 mils the voltage across it must be $150 \times \frac{1}{8} \cdot 7 = 255$ mV. It will be noted, however, that the resistance of the summating resistor associated with this part of the circuit is twice the magnitude of the others, and hence the actual voltage required across the screw-setting potentiometer is 510 mV. The design of the fixed and variable resistance chain controlling the voltage across the screw-setting potentiometers can be evaluated from this figure.

The gauge-setting unit consists of a double-decade potentiometer of total resistance 100 ohms, and a single-decade potentiometer of total resistance 10 ohms. These decades are arranged to give a calibration of 1 mil per ohm step, by adjustment of the 1 000-ohm variable resistor shunting the 42-ohm series resistor.

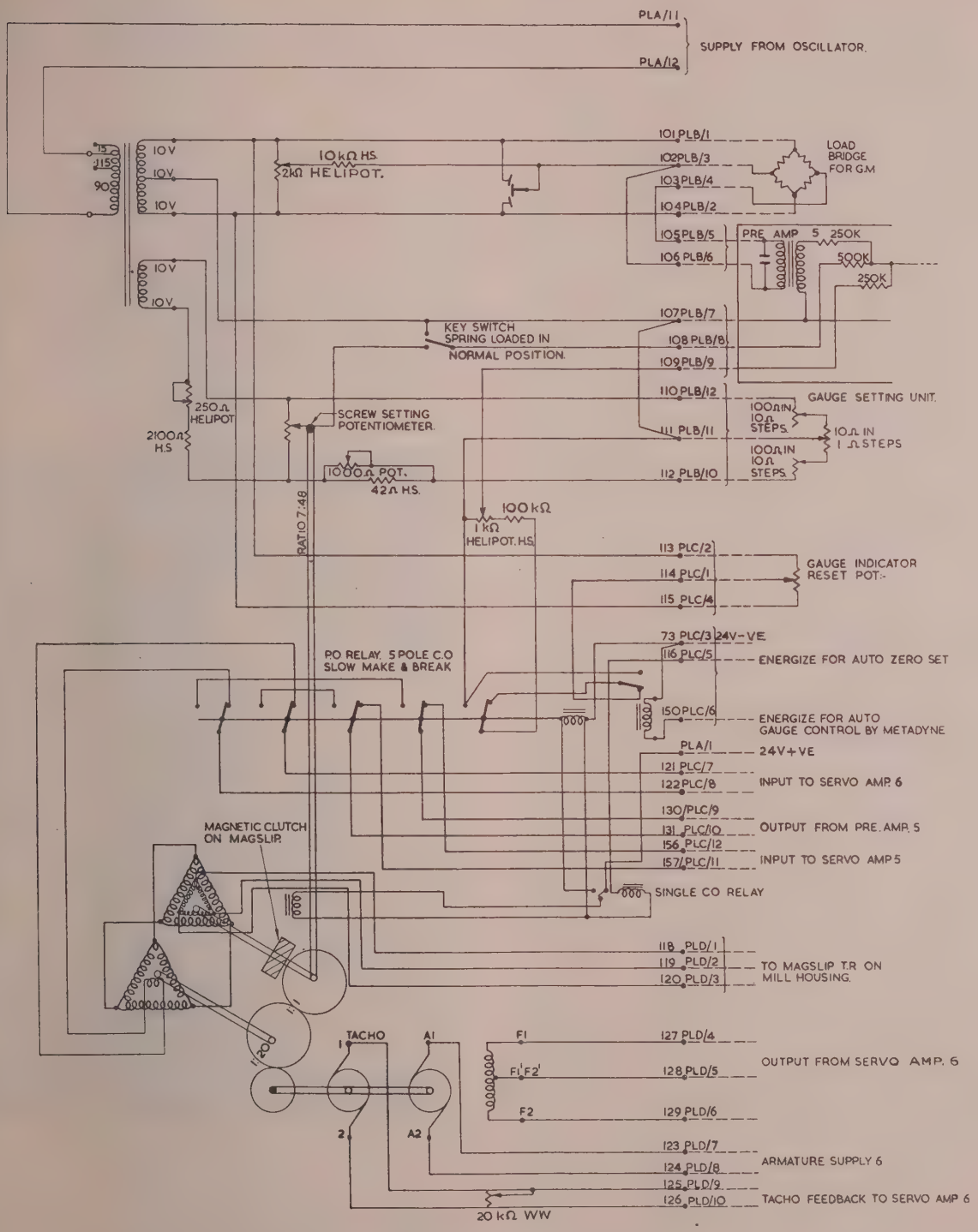


Fig. 11.—Gaugemeter control unit.
The relays are shown in the normal and not the energized position

The fixed 42-ohm resistor and its variable shunt are introduced in the circuit so that when the gauge-setting decade is set at its zero and the rolls are brought into contact and an appreciable load is applied, the screw-setting potentiometer will be capable of being adjusted with its wiper near one end, so that the output signal from this circuit, added to the roll-force meter signal, results in a zero signal on the gauge-deviation indicator.

The spring-loaded key switch is included in the circuit so that the gauge-setting and screw-setting signal circuits can be cut out, and this enables the load bridge zero to be precisely set.

It has been stated above that the screw-setting potentiometer is servo-operated from a Magslip transmitter on the mill housing. This employs the well-known circuit technique of electrically connecting the Magslip transmitter to a resetter and feeding the signal from the rotor of the resetter to a servo amplifier which drives a servo motor geared to this resetter. The servo motor maintains the resetter in angular alignment with the transmitter. The screw-setting potentiometer geared to the resetter thus accurately follows movements of the mill screw. The gear ratios are shown in the Figure and have been chosen to give a satisfactory speed of following.

A system has been introduced which automatically makes the zero of the gaugemeter coincide with the zero of the gauge-setting decades when the rolls are brought into contact. This is done by connecting a follow-through transmitter in the Magslip circuit so that the alignment between the mill screw and the screw-setting potentiometer may be adjusted by rotation of the follow-through transmitter. Thus, by movement of the follow-through transmitter a zero-gauge deviation signal may be obtained under the specified conditions. Details of operation of the automatic zero-set system are as follows:

Assuming that the gauge-setting decades have been adjusted to zero, and the rolls brought into contact with a roll force above the critical value, the zero-set pushbutton is then pressed. This pushbutton is arranged to connect a 24-volt supply on to terminal 116 in Fig. 11; this energizes the change-over relay and actuates the magnetic clutch which now couples the follow-through transmitter to the Magslip resetter with a ratio 1 : 1. The follow-through transmitter is so connected into the circuit that when it is moved in synchronism with the resetter the electrical alignment of the Magslip system is not disturbed, i.e. the resetter output voltage is maintained at zero. Thus under these conditions the servo mechanism can be used to drive the screw-setting potentiometer without disturbing the electrical alignment of the system with the transmitter on the mill housing. The circuit shows that operation of the change-over relay also energizes a 5-pole change-over relay so as to connect the servo amplifier 6, which normally actuates the mill screw servo mechanism, to the output of the gauge-deviation indicator pre-amplifier (pre-amp. 5). This relay also reduces to zero the voltage from the resetter potentiometer in the gauge-deviation indicator. The servo mechanism driving the mill screw-setting potentiometer now operates this potentiometer until the gauge-deviation signal is zero. Thus when the automatic zero-set pushbutton is released, the gauge-deviation indicator reads zero for the particular conditions set before the button was pressed.

The 5-pole change-over relay is arranged to give a slow make and break so that the magnetic clutch operates before the servo-amplifier circuits are broken. This prevents any disturbance of the electrical alignment of the Magslip circuit during the change-over period.

The magnetic clutch was designed by the Admiralty for fitting on Magslip elements. On energization, the Magslip is locked and the gear fitted on it moves freely. This is the normal condition in the system described. On de-energization the Magslip rotor is free and firmly coupled to the gear, so that it can be driven.

This automatic zero-set system has been found very useful in service, since it avoids tedious manipulation of controls to bring the gauge-deviation-pointer indicator to zero, and in addition, the circuit performs the required function and requires no extra equipment, except for a magnetic clutch.

The gaugemeter control unit is built on a 3-unit wide servo-mechanism tray and fits into the electronic rack like the other units.

(3.8) Potentiometric Indication Units

The potentiometric indication units consist of a servo motor which drives through a gear ratio of 4 : 1 down, an indicating pointer and a resetting potentiometer. The latter is a two-gang 3 in-diameter cam-corrected unit. The second half of the potentiometer is used for remote indication or recording the gaugemeter indication. A Magslip transmitter element is also coupled to the resetter potentiometer to provide an alternative system of remote indication. The servo motor is a combined motor-generator unit to provide a feedback signal for stabilization of the servo. These are the essentials of each of the indicator units. The five indicator units are mounted side by side in a long rectangular box and appear from the front as shown in Fig. 12.

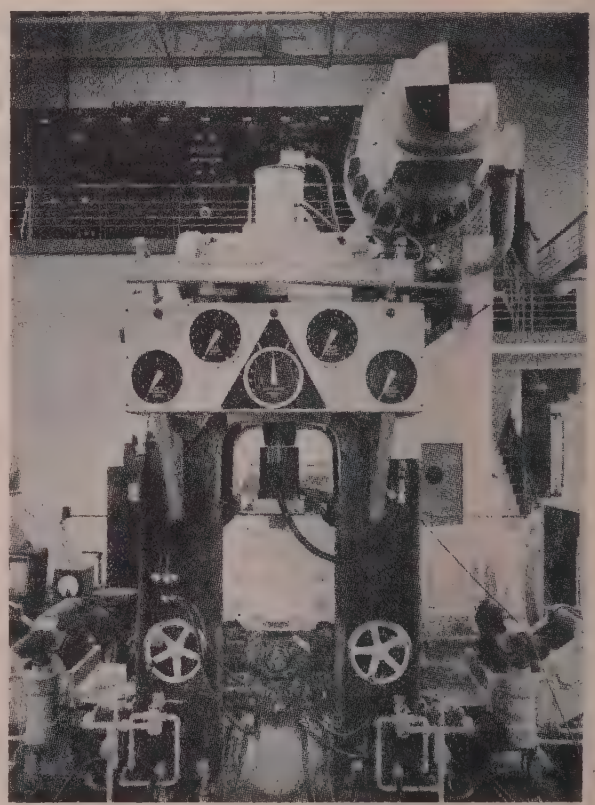


Fig. 12.—Instrumentation system on the 14 in four-high cold reduction mill in the B.I.S.R.A. Sheffield laboratories.

Velocity feedback using the tacho-generator signal has been used for stabilization, as it is usually more satisfactory than phase-advance or equivalent corrective networks in the amplifier where high electronic gains are employed, and consequently the noise level may be high. It also has the advantage that larger tolerances can be allowed in the precision of meshing of the gears in the resetter chain. The use of a low gear ratio between motor and resetter gives the system a fast response. Full-scale deflection is

effected in approximately 0·3 sec with the feedback adjusted to give critical damping.

The gaugemeter scale is calibrated for the range ±0·005 in, the loadmeter scales for 0–150 tons, and the tension meters for 0–5 tons.

(3.9) General Arrangement of the Instrumentation System

All the electronic units and the control units are mounted in the rack, which can accommodate 36 units each of 3 in width. Altogether 30 spaces are used, including those for a spare pre-amplifier, servo amplifier and oscillator. The rack has a terminal compartment which acts as a central junction box for the whole system. Into it are wired all the connections from the 5-channel indicator unit on the mill housing, from the various measuring elements, and from the control elements on the main mill control desk. It also includes relays required in the automatic control systems described later.

On the main control desk is fitted the nominal gauge-setting control and a pushbutton to actuate the automatic zero-set mechanism. Also fitted on this part of the control desk is a change-over switch to introduce automatic gauge control.

The cable runs between the electronic unit and other parts of the system are approximately 50 yd long. The only leads which have been screened are the output leads from the roll-force and coiler torque-measuring bridges. The need for screening has been largely eliminated by careful grouping of all the cables and cores so that the inductance of low-signal-level circuits is small.

(4) PERFORMANCE OF THE MEASURING SYSTEM

The performance of the system is judged on its accuracy, the presentation of the measurement and the speed of response of the indicator. The latter two are determined by the design of the automatic potentiometer system. The automatic potentiometer is also one factor determining the overall accuracy of the measuring system, and since it is common to each measurement, its properties will be considered first, followed by a discussion of the limitations of accuracy which are peculiar to the measurement of roll force, strip tension and gauge.

(4.1) Automatic Potentiometer

The accuracy of the automatic potentiometer display depends basically on the linearity of the potentiometer winding itself, which has a specific linear accuracy of 0·1%. The fixed resistors in the potential dividers of the control units and in the summating circuit of the pre-amplifier affect the calibration, so that their stability affects the accuracy of the system. These resistors can be relied upon to maintain their values to within 0·1%. The loop gain of the servo mechanism is sufficiently high to ensure that the motor will turn for an unbalance of 0·1% of the full-scale deflection. The a.c. and d.c. amplifier gains are chosen so that the drift of the d.c. stages in the servo amplifier produce negligible error. In all, an accuracy of 0·2% is expected from the automatic potentiometer.

The clarity of presentation of the measurement given by the large pointer and dial is an important advantage which may often justify the use of an automatic potentiometer even when its high accuracy is not called for. It is of vital importance in an industrial installation and encourages plant operators to make the fullest use of their instruments. It has already been observed that in this new experimental mill much more attention is paid to the instrument readings in normal rolling than was the case with other experimental mills equipped with small moving-coil indicating meters.

The speed of response is approximately 0·3 sec for full-scale deflection, which is adequate for visual observation.

(4.2) Loadmeters

The accuracy of the roll-force measurement depends primarily on the "load cell" and its strain gauges. There are three main sources of error:

(*a*) *Stress Distribution in the Cylinder.*—There is some departure from linearity in the relations between roll force and surface strain because the distribution of stress over the cross-section of the cylinder varies slightly with the roll force. The relation may also be influenced by off-axial loading on the cylinder, so that care must be taken when calibrating the loadmeter to ensure that the conditions of application of the load are the same as when it is in use. Errors from these sources may amount to as much as 0·8% with the present load cells, but it is thought that this can be improved.

(*b*) *Temperature Effects.*—The strain-gauge bridge is formed from eight gauge elements mounted on the cylinder surface and arranged to measure both axial (compressive) and circumferential (tensile) strains. The arrangement of the elements in the bridge arms is such that effects of changes in temperature are compensated, if it is assumed that all gauges are identical. There is, however, still a possibility of temperature changes unbalancing the bridge, either because of temperature differences from gauge to gauge, or because the gauges do not all have the same temperature coefficient of resistance. The foil gauge has been found to be rather poor in the latter respect, but with wire strain gauges, carefully arranged on the cylinder so as to reduce the possibility of temperature differences, the zero changes with heating are very small. The change with uniform heating of the cylinder is less than 0·02% of full scale per degree centigrade. It has also been observed that a temperature difference of 5°C between points just above and below the gauges, owing to heating of the cylinder from the roll chocks, caused a zero change of only 0·2%.

(*c*) *Creep in the Gauge Adhesive.*—Creep in the gauge adhesive is a well-known phenomenon in strain-gauge work, but when using the modern thermosetting resins its effect is small, and amounts only to 0·2% in an extreme case.

There is also a limit to the accuracy of calibration. The loadmeters were calibrated against a substandard strain-gauge load cell which had been calibrated against the national standard. The national standard for forces above 50 tons is a parallel combination of strain-gauge load cells, each of which has been calibrated against the 50-ton standard deadweight machine.

The accuracy of the standard is given as 0·25%, and the substandard is considered to have been accurate to 0·5%.

The overall accuracy of the experimental mill loadmeter is taken to be ±1%, and the reproducibility better than ±½%.

(4.3) Tension Meters

The tension meters, which are basically dependent on shaft torque measurement using resistance-strain-gauge elements, are limited in accuracy by the same sort of factors as the loadmeters. They were calibrated against a proving ring previously calibrated against a dead-weight machine, the accuracy of the calibration being ±½%.

As constructed, the accuracy of the tension meter is also limited by the radius factor potentiometer, which introduces an error of 2% because of its high resistance. When this is changed, the overall accuracy will be better than ±1%.

(4.4) The Gaugemeter

It was assumed, when describing the principle of the gaugemeter, that the gap between the mill rolls, and hence the strip thickness, was related to the screw setting S and to the roll separating force F according to the linear relation $h = S + F/M$. In fact, this is not exactly true, and the deviations from it are the main cause of inaccuracy in the gaugemeter.

The most important of these errors which arise in the rolling mill itself is that due to thermal expansion of the mill rolls. Despite water cooling, the rolls become hotter than the mill frame and the resultant differential expansion causes a change in the roll gap which is not included in the measurement. In effect, there is an error in the measurement of the screw setting S, its magnitude being approximately 0·000 3 in/°C roll-temperature change. A zero check of the gaugemeter, as described earlier, can be employed at intervals to correct the error, so that the maximum error depends on the frequency of the zero checks in relation to the rate of temperature change. It is impossible to carry out a zero check when strip is in the mill, so that the maximum frequency is one for each coil of strip. The magnitude of the error thus depends in a rather complicated way on the particular circumstances—the type of mill and the work it is doing, the efficiency of the roll cooling, and the size of the coil, and so on. Trials of a gaugemeter system on a large hot mill showed that expansion is not then a serious source of error, but for precise work on the experimental mill it is thought that it will sometimes be necessary to provide means of checking the gaugemeter more frequently, whilst the strip is in the mill, by comparison against some other gauge. An investigation into means of reducing the roll expansion error is in progress.

Another deviation from the theoretical gaugemeter equation is in the relation between mill extension and roll force, which is not strictly linear. The curvature is marked only at low values of roll force, but it exists over the whole range. It is found that for the normal operating range, above 50-ton roll force, the error introduced in this way is less than $\pm$0·000 5 in.

A small correction has to be applied when narrow strip is rolled at high roll loads. This is due to indentation of the rolls where they are in contact with the strip, so that they deform in a slightly different manner from when they are in contact with each other. This results in a slight change of the mill elastic constant which again introduces an error by making the gaugemeter equation incorrect. Instead of applying a correction, it would be possible to provide a control to set the gaugemeter constants according to the strip width.

The present gaugemeter will not give the average thickness of strip with a wedge-shaped cross-section. It will, in fact, either read slightly more than the thicker edge or less than the thinner one, because it measures the roll gap under one screw only. This error, which may be of the order of 0·001 in, can be completely eliminated by taking the mean of the roll gap under the two screws, i.e. the mean of the roll force and the mean screw motion. A simple modification to do this is being made. Two load bridges, one on each side of the mill, will be connected in parallel to give the mean roll force. A follow-through Magslip transmitter on the second screw will, in conjunction with the existing transmitter, give the sum of the two screw motions. This can be halved, thus producing the mean, by altering the gearing in the gaugemeter control circuit by a factor of two.

The pure instrumental errors are, in general, negligible compared with those just described which originate in the mill itself. The mill-screw follow-up servo mechanism, the potentiometer it drives and the remainder of the gaugemeter electrical system have been designed to have a static accuracy of 0·000 1 in. The only other significant source of error is the loadmeter; although its absolute accuracy is of no consequence in this matter, good linearity and reproducibility are essential. As the elastic constant of the experimental mill is about 3 000 tons/in an error of 1·5 tons (1% of full scale) in each loadmeter would result in a gaugemeter error of 0·001 in. The maximum error expected from this cause is 0·000 5 in.

It is concluded that, apart from the thermal-expansion error, the nature of which is discussed fully above, the gaugemeter is accurate to the order of $\pm$0·001 in. This is with the restrictions that the roll force must be above the 50-ton minimum (below which the mill-extension/roll-force relation departs appreciably from the linear) and that a correction is made if the width of the strip rolled is changed.

On some industrial mills where rolling schedules are such that the roll-temperature changes would cause unacceptable errors, it may be necessary to use another gauge-measuring system, such as a flying micrometer or X-ray gauge, to monitor and apply correction to the gaugemeter. Such a procedure would be justified for many automatic gauge-control installations where the high speed of response, which only the gaugemeter system can give, may be essential.

(5) AUTOMATIC GAUGE CONTROL

The instrumentation system has been designed to provide facilities for experiments with a variety of automatic gauge-control systems previously developed by Hessenberg and Sims.[6] Already a simple method of on/off gauge control has been applied, using the gauge-error indicator to provide the control action.

(5.1) Automatic On/Off Gauge Control

In the indicator unit a pair of cams is fixed to the pointer shaft and arranged so as to operate one or other of a pair of microswitches if the gauge deviates by more than a set tolerance from the nominal value on the gauge-setting control. This tolerance zone can be varied by altering the cam settings. When switched to automatic gauge control, these microswitches actuate relays located in the electronic rack, which in turn energize the appropriate screw-motor contactor, thus raising or lowering the screws to provide the necessary correction to the gauge.

The screw drives are provided by a.c. induction motors, and in order to provide quick braking and to minimize overshoot when the control contactors are open-circuited, a technique of d.c. injection has been employed to provide heavy electrodynamic braking. The motor contactors are arranged so that, when the a.c. excitation is removed from the stator, d.c. excitation is applied in its place thus setting up a stationary magnetic field which induces the dynamic braking.

As applied to the mill, this technique of on/off control is at present in its most elementary form. Stable operation of the system has been obtained with a gauge-error tolerance zone as low as $\pm$0·000 5 in. To reduce this further and still maintain stability, it will be necessary to shorten the contactor operating times. When this has been done, still further improvement may be possible by applying a technique of velocity feedback stabilization which has been used in on/off control servo mechanisms of this nature. In this instance a signal would be obtained proportional to the rate of screwing, and injected as a bias into the gauge-error indicator circuit. Thus, assuming that the motors are screwing down to restore a required gauge after a disturbance, the rate signal would cause the gaugemeter to show a gauge within the tolerance in anticipation of that actual gauge being achieved. This will cause the screw motors to be de-energized a short time before achievement of the required gauge, in order to allow for the overshoot due to motor inertia. It must be pointed out that a correcting signal proportional to the rate of screwing is not the theoretical ideal required for complete elimination of overshoot in this instance, but it is suggested as a correction which can be readily applied and is better than nothing. If it is assumed that the braking of the motor on de-energization is entirely electrodynamic, thus producing a braking force approximately proportional to speed, the final resting position of the motor will be approached exponentially and the overshoot will be proportional to the initial velocity. If, on the other hand, the braking of the

motor was entirely due to a fixed frictional force, the overshoot, after de-energization, would be proportional to the velocity squared. Thus a compromise between velocity and velocity squared is required for the correcting signal. Furthermore, the correcting signal should be derived from rate of change of gauge and not from the screwing rate. The relationship between these two is variable and dependent on the rolling conditions, and since gauge error initiates the control action, rate of change of gauge should be used for correction. If, however, a rate-of-change-of-gauge signal were to be injected into the gaugemeter circuit in the required direction, it would have the effect of unstabilizing the gauge-error-indicating servo mechanism. For this reason it is suggested that a screwing-rate signal be accepted, and this signal is already available from the tacho-generator in the servo mechanism positioning the screw-setting potentiometer. It is thus necessary to inject only a part of this signal into the d.c. stage of the gaugemeter servo amplifier to introduce the required velocity-feedback stabilization.

(5.2) Automatic Gauge Control by Continuous Action

By the use of the equipment available, a system of automatic gauge control can be operated in which a correction to the screws, or other devices affecting the gauge, can be applied which is proportional to the gauge error. To do this the servo amplifier, at present actuating the gauge-error indicator, would be switched to the appropriate control field of a metadyne or Ward Leonard set for screw position control or to an electrohydraulic valve in an hydraulic control system. Relays have already been wired up in the electronic rack to effect this change over. With this system, however, there would be no indication of gauge error when operating under automatic gauge control. In order to provide a simultaneous gauge-error indication it will be necessary to add an extra amplifier channel consisting of a pre-amplifier and servo amplifier. These amplifiers would be used to generate the error signal to actuate the automatic gauge control leaving the gauge-error indication unaffected. The inputs to the additional pre-amplifier would be from the roll-force bridge output and gauge-setting control circuit only, thus excluding the signal from the resetter potentiometer.

(6) AUTOMATIC TENSION CONTROL

Facilities have been provided for automatic tension control based on measured tension instead of by the present system of control of coiler-motor armature current with correction for radius which gives only an approximate automatic tension control system.

The method provided uses the tension indicator unit with its potentiometer to set the required tension, and the output of the servo amplifier normally associated with the tension indicator is switched over to the control winding of the machine determining the coiler tension. Thus any deviation of tension from the value set on the indicator gives rise to an error signal actuating the servo amplifier and so generates the necessary control action. The switchover of the servo-amplifier output is effected by a relay which is already wired up in the electronic rack. In order to set the pointer of the tension indicator unit at the particular tension set by a knob on the control desk, an M-type step transmitter is used. This is a step transmission system using d.c. energized rotary transmitter and receiver elements. An M transmitter on the control desk is energized when the system is switched to automatic control. Also a magnetic clutch in the indicator unit is made which couples an M receiver to the pointer shaft, thus permitting its adjustment remotely from the desk. The components of this remote-control system are available and await fitting.

There is an alternative system of automatic tension control which may be slightly less convenient in operation but does not require a remote transmission system. In this, the tension is indicated normally and brought up to the required value under the existing manual control. The output of the servo amplifier operating the indicator unit can be switched to the control field of the machine determining tension, and thereafter the tension will be held at the set value. The operational inconvenience of this system is that to change the tension setting it is necessary to revert to manual control and then back to automatic control. The system has, however, the advantage that the amount of tension adjustment to be made through the winding of the machine controlling tension is small compared with the first method described, which has to give sufficient control action to operate over the whole range of tension. It is therefore likely that this second system will be more suitable for the existing electrical control system.

(7) CONCLUSIONS

The system of measurement and display of roll loads, strip tension and strip gauge described in the paper has shown itself very satisfactory in service and has served as a useful model upon which industrial instrumentation schemes might be based. The rationalization of the electronic units and the choice of a potentiometric system of display have helped greatly in getting this comprehensive scheme quickly into service.

Experiments so far carried out with a simple system of automatic gauge control by on/off control of the screw motors, using the gauge-error indicator to initiate the control action, have shown that this system has considerable promise for use on some industrial mills. The ultimate performance of the automatic gauge control system will, however, be dependent on the condition of the mill to which it is applied and particularly on roll eccentricity.

(8) ACKNOWLEDGMENTS

The instrumentation system described in this paper was partly developed and wholly designed, built, installed and adjusted in under three months. In achieving this object the authors wish to acknowledge the special efforts of their colleagues in this work, notably Messrs. J. Manuel, F. D. Grimmer and J. R. Spencer of the Instruments Section; Mr. A. Hambly and his workshop staff, and Mr. R. M. J. Garnett who was responsible for a large part of the mechanical design; also to members of the Rolling Section for their assistance in installation. Acknowledgments are also due to Ferranti Ltd. for quick delivery of essential electronic equipment employed. Acknowledgment is also due to Mr. R. F. Bowler, who was concerned with much of the early work on mill-measuring systems.

(9) REFERENCES

(1) British Patents Nos. 681373 and 692267.
(2) Sims, R. B., Place, J. A., and Morley, A. D.: "Loadmeter for Industrial Mills," *ibid.*, 1951, **173**, pp. 116 and 137.
(3) Rankine, J., Bailey, W. H., and Stanton, R. P.: "Resistance Strain Gauges for the Measurement of Roll Force, Torque and Strip Tension," *Journal of the Iron and Steel Institute*, 1948, **160**, p. 381.
(4) Sims, R. B.: "Gaugemeter for Strip Mills," *Engineering*, 1953, **175**, p. 33.
(5) Sims, R. B.: "The Design of a Tensiometer," *Engineering*, 1952, **174**, p. 232.
(6) Hessenberg, W. C. F., and Sims, R. B.: "Principles of Continuous Gauge Control in Sheet and Strip Rolling," *Proceedings of The Institution of Mechanical Engineers* (A), 1952, **166**, p. 75.

[The discussion on the above paper will be found overleaf.]

DISCUSSION BEFORE THE MEASUREMENT AND CONTROL AND THE UTILIZATION SECTIONS, 6TH DECEMBER, 1955

Mr. W. Spence: I feel that the paper summarizes a great deal of useful work towards the closer control of the product of cold strip mills. It also represents a very bold policy on the part of the authors in proposing the use of such a plethora of electronic equipment. While the steel industry is beginning to accept the use of electronic equipment where it is obviously essential, the tendency is still to put complete reliability first and to regard electronics with suspicion from this point of view. This is probably the main reason for the enthusiastic reception of magnetic amplifiers by the steel industry.

I should like to make a plea to those concerned with the specification of new equipment, however, not to proceed exclusively with any one particular type of amplifier, either rotating, magnetic or electronic. Each of these amplifiers has its advantages, but at the same time they all have their disadvantages. The engineer who is designing the control system is in the best position to decide which is the most suitable amplifier to use in any particular circumstances. If he has a free hand, the best control system should result.

I endorse the feature of rationalization of the various units so that identical amplifiers are used. This procedure, coupled with a satisfactory mechanical design which permits rapid replacement of a faulty unit by a spare, removes most of the stigma of complexity and vulnerability from electronic equipment and makes maintenance a reasonably straightforward proposition.

The term 'resetter', for a monitoring device providing the main feedback signal in a regulating loop, has been used in a number of different ways in the past, and I suggest that it would probably be best to allow it to fall into disuse. It has been deprecated by B.S.I. committees on nomenclature.

In Section 3.7, there is a description of a Magslip-controlled servo mechanism, coupling the screw-setting potentiometer to the screw. A possible disadvantage of this is that a further resilient linkage or time delay has been introduced into the system. What inaccuracy or difficulty in stabilization has thus been introduced?

Has experience with plugs and sockets for the electronic units been completely satisfactory?

The radius-correcting potentiometer described in Section 3.6 would hardly be acceptable on a production mill. Has any other means of obtaining a measure of coil diameter been tried?

Finally, I should like to know the reasons for the choice of the M-type transmission system in the tension-setting arrangement mentioned in Section 6.

Mr. W. N. Jenkins: I agree that the scheme of instrumentation described is a model for industrial mills of the same class, and I think that other types of mills would be improved by the application of the techniques described in the paper.

I make a reservation, however, in connection with the method of radius measurement. The radius-correcting potentiometer would cause difficulties in many production mills and would be liable to damage by cobbles, which do occur from time to time. There is an alternative method of radius measurement which would probably be accurate enough for industrial mills and which would be possible with most modern coilers. The control of such coilers is by the maintenance of constant current in the motor armature and by matching the motor back-e.m.f. to the strip speed. In such schemes, the coiler field current is automatically maintained proportional to the coiler radius.

The error mentioned in the paper of $\pm\frac{1}{2}$ mil due to the non-linear relationship between extension and roll force is quite a significant percentage of the thickness of the lighter gauges rolled. It is an error which would appear as a gauge error when automatic gauge control is applied through the screws, particularly when the ingoing gauge varies. It would be interesting to know by how much a given fractional change of input gauge is attenuated by the experimental mill, both with and without the application of automatic gauge control.

Mr. J. B. C. Robinson: How far could these particular systems be regarded as suitable for use in production mills? It sometimes rather alarms me to observe how electronic equipment is installed apparently ruthlessly on some occasions, and sometimes it is not too well designed.

I suggest that when considering electronic equipment of this complexity, alternative methods of control should be provided, so that the mill can operate, albeit with some reduced efficiency, and the maintenance staff can have a chance to correct any faults that may arise.

I do know of one tension-control system at least, working through feedback systems on a tension meter, which actually uses no electronic equipment whatsoever, but employs an all-magnetic amplifier system. Thus, for those who are not too happy with electronic equipment, it will be seen that there are alternative systems which offer a little more robust reliability, such as is needed in the rolling mill.

How long does it take to set up this type of equipment, assuming, for example, that a unit which has been in service for a long time is replaced by another which requires setting up? There seem to be rather a large number of independent adjustments involved.

All equipment, whether it uses electronic apparatus or magnetic amplifiers, on occasion suffers from inadvertent failure. In some cases these failures have resulted in very serious and very expensive accidents.

Mr. J. F. Coales: There is a great future for on–off controls for the reason given by the authors, i.e. they save the motor from running continuously for small errors which are, in fact, of no importance, provided that the on–off control has a dead zone.

In this case, I imagine that a sufficient error must be allowed, so that the periods when the gaugemeter shows an excessive error which has to be corrected are only a very small part of the time of operation. This principle could be employed in many more industrial operations than at present, largely because a fetish for linear controls has arisen.

With regard to the change of roll diameter due to temperature, have the authors considered the possibility of feeding in a temperature correction? Presumably, the temperature of the rolls can be fairly easily measured. I know there are some difficulties, because it will not be even throughout the roll, but, if this could be done, it could then be used as an open-loop control, varying the input to the control system. This might overcome some of the difficulties experienced with multi-loop controls.

The reliability of electronic equipment has been questioned, and there have indeed been many poor designs of electronic equipment, but I do not think there is any need to have unreliable electronic equipment in these days. The life of electronic valves can be very long indeed, and experience with equipment using thousands of valves shows that, if proper precautions are taken and proper maintenance schedules are observed, there need not be catastrophic failures. However, they have a definite life, but I am assured that junction transistors are now available which will fulfil all the roles that are required for this type of control equipment in industry. The next few years should see a combination of transistors and magnetic

amplifiers which should provide very much increased reliability in control equipment.

Dr. Denis Taylor: Thermionic-valve circuits can be extremely reliable when properly designed with the right components, but this is not always appreciated. I remember some years ago that there was criticism on the score of the reliability of the control systems used for nuclear reactors, because thermionic valves were employed. It is possible to design a control system for a nuclear reactor without using any thermionic valves, and I have described such a system.* However, they have not been used, because the thermionic-valve systems are more convenient and, if properly designed, can be extremely reliable.

I have not examined in detail the circuits described in the paper, but it is evident that magnetic amplifiers could be used to replace the thermionic-valve amplifiers in at least some places; this would probably be worth while. There has been some discussion about replacing the thermionic valves by transistors, and it would be interesting to know whether any improvement in reliability would result from this change.

Mr. G. Syke: I consider that too much emphasis is being placed on the high speed of response of the measuring element. I think that the ultimate limitation to speed of automatic control will generally be set by the dynamic characteristics of the mill and its drive; almost any type of measuring element one would care to consider for the purpose—certainly the right sort of radiation thickness gauge—is capable of keeping pace with this.

We have heard of inevitable lag in the control system due to inertia, friction and backlash associated with the screw. Is it not therefore preferable to measure some way back before the rolls, and thus anticipate any sudden thickness changes?

On reversing mills it is usual to have a thickness gauge on either side of the mill. This arrangement lends itself to obtaining the higher-frequency a.c. components of the measured variable from the entry-side gauge and the lower-frequency or d.c. components from the exit-side gauge. The first anticipates and the second checks the results.

The experience gained in reliable electronic nuclear-reactor instrumentation is being liberally passed on by Dr. Denis Taylor and his staff to industrial instrument designers; we try to make good use of this and are most grateful. On rolling-mill installations the electronic equipment is usually in the control room under favourable conditions, and a high standard of reliability can thus be assured. The measuring elements situated on the mill are subject to mechanical damage, and they work under bad conditions in respect of oil, dust and temperature. These usually present the more difficult design problems, and are less readily accessible for maintenance than the electronic equipment.

Mr. J. I. Bernard: I know that those users with most experience are quite satisfied that electronic control is reliable. Admittedly this applies to the larger organizations who no doubt have staff trained for any necessary maintenance work. Doubts seem to arise almost entirely with people who have had no experience of it. This bears out Dr. Denis Taylor's comments.

Provided that the gear is well made and the firms making it are sufficiently experienced in its design to know this, there is nothing to be feared from the use of electronic control. The increased reliability and precision that come from the much greater magnification and accuracy obtained with such systems of control are most certainly very helpful in increasing productivity and are a real help in solving the economic difficulties of the present day.

Mr. R. B. Sims (*communicated*): I know well how much thought has been given to the design of this instrumentation. The layout, however, seems to be very inflexible, and roll torque is not included. Moreover, the equipment is costly and could not be used for recording without further considerable expenditure. In my experience of research, the only way to collect the data reliably was to use high-accuracy, high-speed recordings, the information being collected on moving-coil meters and photographed on 35 mm film. This, when recording up to 20 channels of information, removes the time error. It is chiefly on a mill in production that the system described by the authors shows a marked superiority; another advantage which they may fairly claim is the higher accuracy of potentiometric methods.

I do not think that the method of tension measurement described will ever be used again. It would be unacceptable in production mills, and the 3-roll tension-meter installed on the pass-line is better in every way. Automatic tension control using a tension-meter* promises to be simpler electrically and more accurate than the constant-current tension-control systems.

A load-meter, uniformly loaded, should be reproducible over a very long time at the 0·1% level or better, and the temperature drift of 0·02% per deg. C, found by the authors, is greater than normal. The errors in load-meters, which can be larger than the authors describe, can be eliminated by good design. The effect of adhesive creep is negligible, as Bowler† has shown.

I should like to know whether the authors are confident that they could obtain an accuracy within ±0·2% from their electronic circuits and hardware when working with their designs of resistance strain-gauge networks, with normal mains-supply fluctuations and under all operating conditions.

The errors in a gauge-meter are still a matter for discussion, but, in production mills, thermal effects are not likely to be serious. The first industrial automatic-gauge-control unit, which has shown the thermal changes to be almost negligible, is of the on–off type, as described in the paper, but with sundry modifications. It can hold all gauges to ±0·0007 in, and has proved to be a practical industrial unit. An automatic-gauge-control unit operated from Ward Leonard controlled screw motors‡ would, however, be superior to the on–off type; and so, too, would the hydraulic analogue described by Sims and Slack.§

THE AUTHORS' REPLY TO THE ABOVE DISCUSSION

Messrs. S. S. Carlisle and **G. W. Alderton** (*in reply*): We agree with Mr. Spence that the class of amplifier recommended for a particular job, e.g. rotating, magnetic or electronic, should be decided on the merits of the specific application, but there is great scope for a rationalization of design within any one class of amplifier.

No difficulty was experienced due to the dynamic characteristics of the Magslip screw-setting servo link, because the limiting velocity and acceleration of screw setting are well within the range of easy design of low-power servo mechanisms.

Plugs and sockets have been found satisfactory during two years' service on this installation, but for industrial use permanent cable connections may be better.

The radius-correcting potentiometer was not considered entirely satisfactory. Other systems of radius measurement were not tried, however, because the three-roller tensiometer is undoubtedly the better tension-measuring system.

M step-by-step transmission was proposed as a ready means

* Discussion on 'Symposium of Papers on Nuclear Reactor Instrumentation', *Proceedings I.E.E.*, May, 1953 (100, Part I, p. 116).

* Sims, R. B., Place, J. A., and Briggs, P. R. A.: *Journal of the Iron and Stee Institute*, 1953, 173, p. 354.
† Bowler, R. F.: *Instrument Practice*, 1953.
‡ Sims, R. B., and Briggs, P. R. A.: *Sheet Metal Industries*, March, 1954.
§ Sims, R. B., and Slack, K. H.: *Proceedings of the Institution of Mechanical Engineers*, 1955, 169, p. 255.

of giving the torque and accuracy required without the complication of a power-amplifier servo system.

In reply to Mr. Jenkins, the possible error quoted of $\frac{1}{2}$ mil due to non-linearity between roll force and mill extension is a limiting value which will occur only for very large changes in roll force. Such changes would not occur in normal operation of the automatic gauge-control system. We agree that other better methods of coil-radius measurement could be used, but as already stated, tension is best measured by the three-roll tensiometer.

In reply to Mr. Robinson the system we have described is considered suitable for use on some production mills, and one installation is now operating on a production cold mill. Trials are in progress which indicate that the system is satisfactory. In this method of automatic gauge control a return to manual control can be made easily in the case of breakdown of the equipment.

For tension control, magnetic-amplifier-operated systems may be suitable and more reliable. We are firm believers, however, in the reliability, for industrial use, of a well-designed and properly installed electronic amplifier system. It has the advantage over a magnetic amplifier in that it has a better frequency response.

There is not yet sufficient experience on production installations to state the setting-up time required. The system described in the paper could be improved in this respect.

We are pleased to have Mr. Coales's encouraging remarks about the potential uses of on-off control. We have found it simple and advantageous in this case.

Correction for roll-temperature change is difficult, mainly because of the lack of a suitable method of measuring roll-surface temperature. We have not yet solved this problem. Probably the best solution is to be found in the use of a secondary strip-gauge monitoring system as suggested in the paper.

We entirely agree with Mr. Coales that much of the criticism of the reliability of electronic equipment has arisen as a direct result of bad design. It is a great misfortune that it is so easy to assemble and make work an electronic measuring and control system, even though it may be a system of bad design or not even designed at all. Because of this ease in making them work, systems are often prematurely applied in industry.

In reply to Mr. Syke, we would agree that other measuring systems, e.g. a radiation gauge, have an adequate speed of response for use in automatic gauge-control systems applied to present-day mills with simple motor-controlled screws. It is likely, however, that the response of mill-roll loading systems will have to be speeded up, either by improved motor controls, such as Ward Leonard systems, or by further development of hydraulic roll-loading devices. This increase in speed will be necessary to give adequate gauge uniformity at high rolling speeds and mill accelerations. These requirements will in the future, if not now, justify the higher response speed which the B.I.S.R.A. gaugemeter system can give.

It is desirable to measure gauge earlier, and so anticipate changes in finished gauge. Consideration has already been given to this system among many others concerned with the whole problem of gauge control in tandem mills.

We are very glad to have Mr. Sims's contribution and comments, particularly in view of his pioneering work in this subject. It is encouraging to note his statement on the performance now being obtained with the first industrial automatic-gauge-control unit on a production mill.

We would agree that the Ward Leonard control of screw motors is capable of giving superior performance to an on-off control system, but full justification of its extra cost on screw-loaded mills is dependent on further improvement in gauge-measuring systems.

The measuring system we have described was designed to include facilities for recording from the Magslip transmitter elements. Admittedly the method Mr. Sims employed of high-speed milliameters and film recording gives better frequency response and is cheaper. The performance of the potentiometric system is, however, considered adequate for research purposes; it enables paper records to be obtained without processing, and serves the other purpose of giving a clear instrument display to aid mill operators.

We are pleased to note Mr. Sims's confidence in obtaining improved load-metering performance.

We consider an accuracy of $\pm 0{\cdot}2\%$ quite practicable in a potentiometric system of this type with carefully designed circuits and resistive networks. The gains of the servo amplifiers are sufficient to ensure that normal mains-supply variation will not introduce more than $0{\cdot}1\%$ error.

681.142 : 621.317.79

Paper No. 1899 M
Oct. 1955

TRIDAC, A LARGE ANALOGUE COMPUTING MACHINE

By Lt.-Comdr. F. R. J. SPEARMAN, R.N.(Ret.), A.M.I.Mech.E., J. J. GAIT, M.A., B.Sc., A.Inst.P., Associate Members, A. V. HEMINGWAY, B.Sc.(Eng.), Member, and R. W. HYNES, B.Sc.

(*The paper was first received 28th September, in revised form 1st December, 1954, and in final form 19th April, 1955. It was published in October, 1955, and was read before* THE INSTITUTION *1st December, the* NORTH-WESTERN CENTRE *6th December, 1955, the* NORTHERN IRELAND CENTRE *14th February, and the* SOUTH MIDLAND RADIO GROUP *26th March, 1956.*)

SUMMARY

Tridac is the name given to a large analogue computing machine installed at the Royal Aircraft Establishment. The machine, which has been built to assist in solving guided-weapon problems, has electronic, mechanical and hydraulic components, and operates on a 1 : 1 time scale so that real components can be included in the computation. Tridac is intended to assist system understanding and development by constructing system models with which mathematical computations can be carried out. Each computing section, the parameters of which can be easily changed, represents a particular part of the real system.

The mathematical operations of summation, integration, multiplication and resolution are carried out using either drift-corrected d.c. amplifiers or electrically controlled servo motors, with hydraulic or electric power. Connections to the servo motors and interconnections between the various cabinets, which contain in all some 2 000 electronic units, are arranged to facilitate the simulation of the flight of a guided missile.

Principles of operation, computing methods and components, installation, and method of operation of the computer are discussed with reference to the problems to be solved.

(1) INTRODUCTION

Early in the post-war programme of research on automatic control of aircraft and missiles undertaken at the Royal Aircraft Establishment, it became apparent that aids in the form of mathematical machines or simulators were necessary. The first machines developed were small electronic differential analysers based on the well-established principle of using a high-gain d.c. amplifier with known input and feedback impedances arranged to achieve a desired transfer function between the output and input voltages of the unit. The value of such aids to the research programme was quickly established, one of the greatest benefits being the insight obtained into the physical operation of the system under study. From these early beginnings up to the present day, the machines used have been regarded as "aids to physics" or "aids to engineering" rather than "aids to mathematics." They are regarded as working models of the system under study as much as mathematical calculating machines.

Two main lines of development have been followed, one directed towards achieving higher accuracy in the operation of existing units, and the other towards extending the scope of the machines to include more types of operations such as multiplication, the generation of functions of variables, and simulation of trigonometric relationships.

The majority of the analogue computers built at the R.A.E. for the solution of problems associated with the high-speed flight of aircraft and missiles have operated on a "real" (or 1 : 1) time scale. Since the duration of such problems is usually of the order of seconds or minutes there is no real need to speed up events in the simulation, and the use of a unity time scale offers the valuable facility that human operators or actual parts of the system under test may be incorporated into the simulator. Apart from the simulator about to be described, however, all these machines have suffered from certain serious limitations, and their capacity was such that flight in one plane only could be studied.

For studies of controlled 3-dimensional flight with six degrees of freedom, and including aerodynamic cross-coupling effects, kinematic couplings and control and guidance cross-couplings, a computer of very large size and complexity, as used at the Massachusetts Institute of Technology,[1] is required. Since only one such computer could be contemplated at the R.A.E., the machine was specified to be as flexible as possible and to operate on a real time scale in conjunction with real aircraft or missile components. To obtain consistency of results the design had to be such that each of the many individual mathematical operations in the solution of the complete problem had to be performed with high accuracy. Digital methods were, however, not applicable because of the real-time requirement, which, together with the large number of equations to be solved and the speed of response demanded, would have required a prohibitively high digital rate of computation, and necessitated the development of accurate and rapid digital/analogue convertors. Attention was therefore concentrated on the design of an analogue computing machine, or simulator, using standard analogue computing techniques but requiring the development of components of a higher order of accuracy than previously used.

Active work on Tridac,* as the computer is called, began in 1950 subsequent to a visit to the United States to study simulator work there. The development, construction and erection of the computer at R.A.E. was completed in 1954. As can be seen from Fig. 1, this constituted a major engineering effort.

Tridac is basically an analogue computing machine suitable for the calculation in three dimensions of the trajectory of one aircraft pursuing another, or of a guided missile attacking a target. As such it first needs to compute the aerodynamic forces and moments on the pursuing aircraft or missile resulting from automatic or human control, and the resulting angular and linear motions. Secondly, it has to compute the kinematics, or relative motion between pursuer and pursued, and the error signals which, through the automatic or human pilot, control the flight path of the pursuing aircraft or missile. It is so designed that each functional part of the system under test, such as the aerodynamics or the control-surface servo motor, is represented by a self-contained computing block of the machine with an input and output. This, together with the 1 : 1 time scale already mentioned, enables actual components (e.g. control-surface actuators) to be used in computations. Working models of various systems can thus be constructed, with a mixture of real and simulated components.

The units of Tridac perform the standard mathematical

Lt.-Comdr. Spearman is at the Ministry of Supply, and was formerly at the Royal Aircraft Establishment.
Mr. Gait is at the Royal Aircraft Establishment.
Mr. Hemingway is with Tube Investments, Ltd., and was formerly with Elliott Brothers (London), Ltd.
Mr. Hynes is at the Long Range Weapons Establishment, Australia, and was formerly with Elliott Brothers (London), Ltd.

* Tri-dimensional analogue computer.

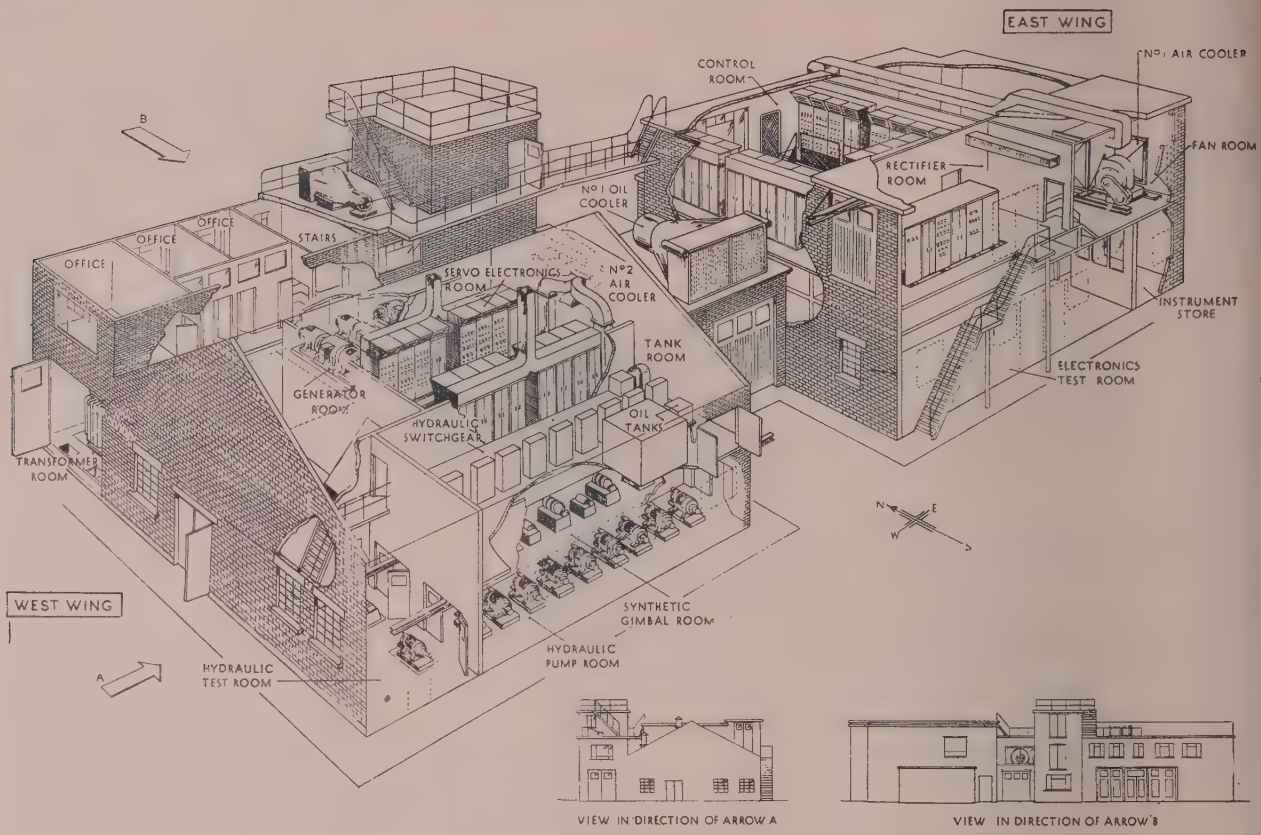


Fig. 1.—General layout of Tridac.

operations with high accuracy, and the method of connecting the computing units together within a computing block, such as that which computes the aerodynamic forces, is flexible, and changes can easily be made. The permanent signal wiring between the various computing blocks is, however, arranged to suit the particular problem of an aircraft or missile pursuing a target. Problems other than this can be solved by means of Tridac, provided that the units performing the different mathematical functions required are available in the correct numbers at suitable locations. Modifications to the permanent signal wiring would, however, be necessary should the simulator be required to solve problems of a different type.

Quantitative results of Tridac computations are recorded on various types of recorders and on plotting tables for subsequent analysis. For qualitative analysis, the results are also displayed on cathode-ray tubes, with a tape recorder for replaying.

(2) DESIGN CONSIDERATIONS

(2.1) Choice of Computing System

Since computing was to be done by analogue methods, the various known techniques of analogue computing were reviewed. A high-speed repetitive type of machine was not suited to real-time-scale operation, and a non-repetitive or "single shot" machine operating on direct-voltage analogue principles was selected in preference to either a mechanical or electro-mechanical machine or to a machine operating on alternating-voltage analogue principles, as offering greater promise of flexibility with adequate accuracy, for the following reasons:

(*a*) Although mechanical systems may offer some increase in accuracy in certain operations, the attainment of this accuracy is very expensive and the resulting machine is less flexible than its electrical counterpart; the desirable 1 : 1 time scale is also generally unobtainable for the high-speed-flight class of problem.

(*b*) Although electro-mechanical servo mechanisms offer some advantage in giving a shaft rotation as an output, thereby facilitating multiplication and resolution, their cost, complication and performance are such that their universal use is not practicable.

(*c*) When the analogy is in terms of direct rather than alternating voltages the operations of integration and differentiation are more readily performed.

(*d*) Direct-voltage units for multiplication, resolution and curve generation had been developed, even though their accuracy was in some cases inadequate for immediate application to Tridac.

The main drawback in a direct-voltage system was its proneness to drift, and it was essential that this be overcome by developing a direct-coupled amplifier with adequate long-term stability. In Tridac the drift has been kept within acceptable limits by associating each d.c. amplifier with a mechanical chopper and a.c. amplifier, or with a magnetic modulator and amplifier. A monitoring system gives a visual warning signal whenever excessive drift, indicative of a fault, occurs in any of the drift-corrected amplifiers.

(2.2) Layout

The layout of Tridac, as shown in Fig. 1, was partly conditioned by the decision to modify existing buildings to house it.

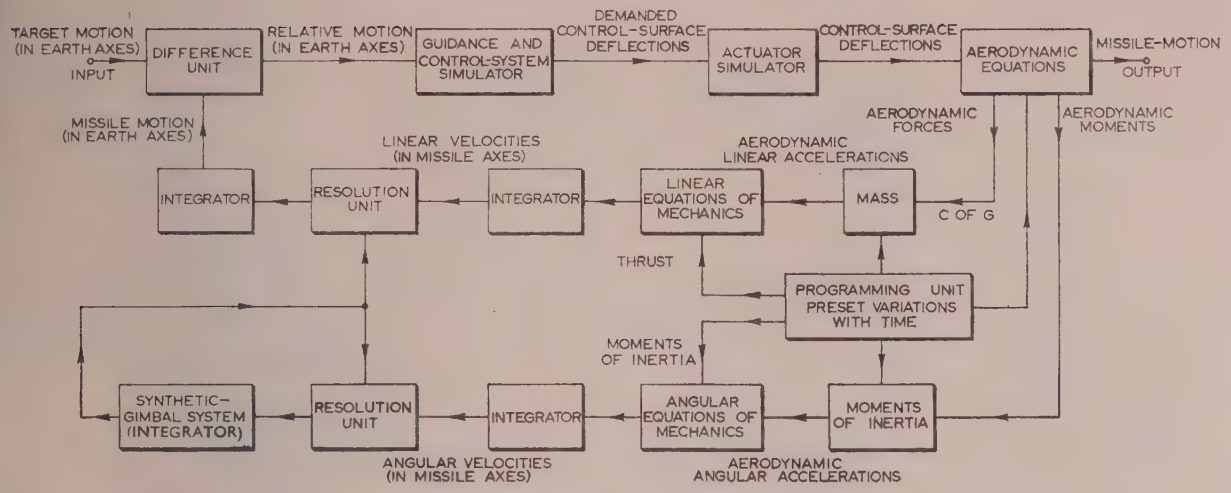


Fig. 2.—Block diagram of Tridac as a closed-loop system.

Major layout design considerations were noise and fire prevention, ease of maintenance, ease of operation, and heat dissipation.

(2.3) Ease of Operation

Ease of operation is an important aspect, and this includes ease of setting up a problem, and indication, location and correction of any faults or overloads during a trial. Facilities for quick examination of the results, means to change a parameter easily, and adequate control over the running of a trial, are also necessary.

Since each block of the simulator corresponds to an actual part of the system, it is easy to vary one block while leaving the others intact. Patch panels are fitted whereby the computing circuits may be changed, and potentiometers allow constants and scale factors to be quickly changed. Removable panels, which enable computing circuits to be set up and stored for future use, are fitted at key points.

For running the trials all controls of the machine are concentrated at the control desk, which has a communication system connecting the machine operator with the other sections of the simulator. Certain faults which may occur are indicated on the control desk, on which are also mounted the cathode-ray-tube displays, which show qualitatively how the run is proceeding.

(3) PRINCIPLES OF OPERATION

(3.1) Application

The main application of Tridac is to the solution of the problem of a guided missile attacking a target in 3-dimensional space, taking into account missile aerodynamics, target motion, autopilot characteristics, the nature of the guidance and control systems, and kinematics of motion. The overall system has several interrelated feedback loops, e.g. the control-surface actuator, the guidance and the kinematic loops.

In dealing with problems of this type the interacting forces and moments (whether in the airframe, aerodynamics, control system or elsewhere) require to be taken into account. The simulator is required to convert such quantities, computed relative to rotating and translating principal body axes of the missile, into quantities referred to fixed non-rotating ground axes or to other moving axes such as those of a radar reflector. Some of the major computations required are thus concerned with axis transformation.

For the particular problem of a missile attacking a moving target the appropriate layout is shown in the simplified block diagram of Fig. 2, which shows Tridac as a closed-loop system, with the target motion as the input and the missile motion as the output. The input can be represented by the target's Cartesian co-ordinates referred to the earth axes, and the quantity fed back can be the missile position similarly represented. The object of the guidance and control system and actuator simulators, working through the missile aerodynamics, is to reduce to zero the error between input and feedback, and hence the input and output, within the time of flight of the missile.

The error is fed to the guidance and control-system simulator, which is set up to demand control-surface deflections in accordance with the navigation law and control system used, e.g. beam riding or homing with proportional navigation.[2] The resulting control-surface deflections come from the actuator simulator, which may be replaced by the real actuator with simulated aerodynamic loading if desired, and are fed to the unit which solves the aerodynamic equations. For a cruciform missile the equations can be expressed in the form given in Appendix 11, and the outputs are the forces along and the moments about the principal body axes of the missile. These aerodynamic forces and moments produce linear and angular accelerations, which depend on the mass and moments of inertia of the missile, and which in turn vary with time and can be preset in the programming unit to vary with time.

The missile motion is computed relative to the translating and rotating axes of the missile and is converted to instantaneously fixed axes by the equations of mechanics given in Appendix 11. The angular motion is used to set up shaft angles which constitute a synthetic gimbal system. This computes the axis transformation matrices as given in Appendix 11, and the matrices are used to resolve the missile linear motion from missile axes to earth axes, so that the difference with respect to the target motion may be obtained.

(3.2) Block Schematic of Overall System

The block schematic of the overall system is given in Fig. 3, in which the various computing blocks, which are interconnected by permanent signal cables to comprise the simulator, are shown with their electrical and mechanical inputs and outputs. A brief description of the block schematic is given, starting at the point in the loop where there are missile control-surface deflections.

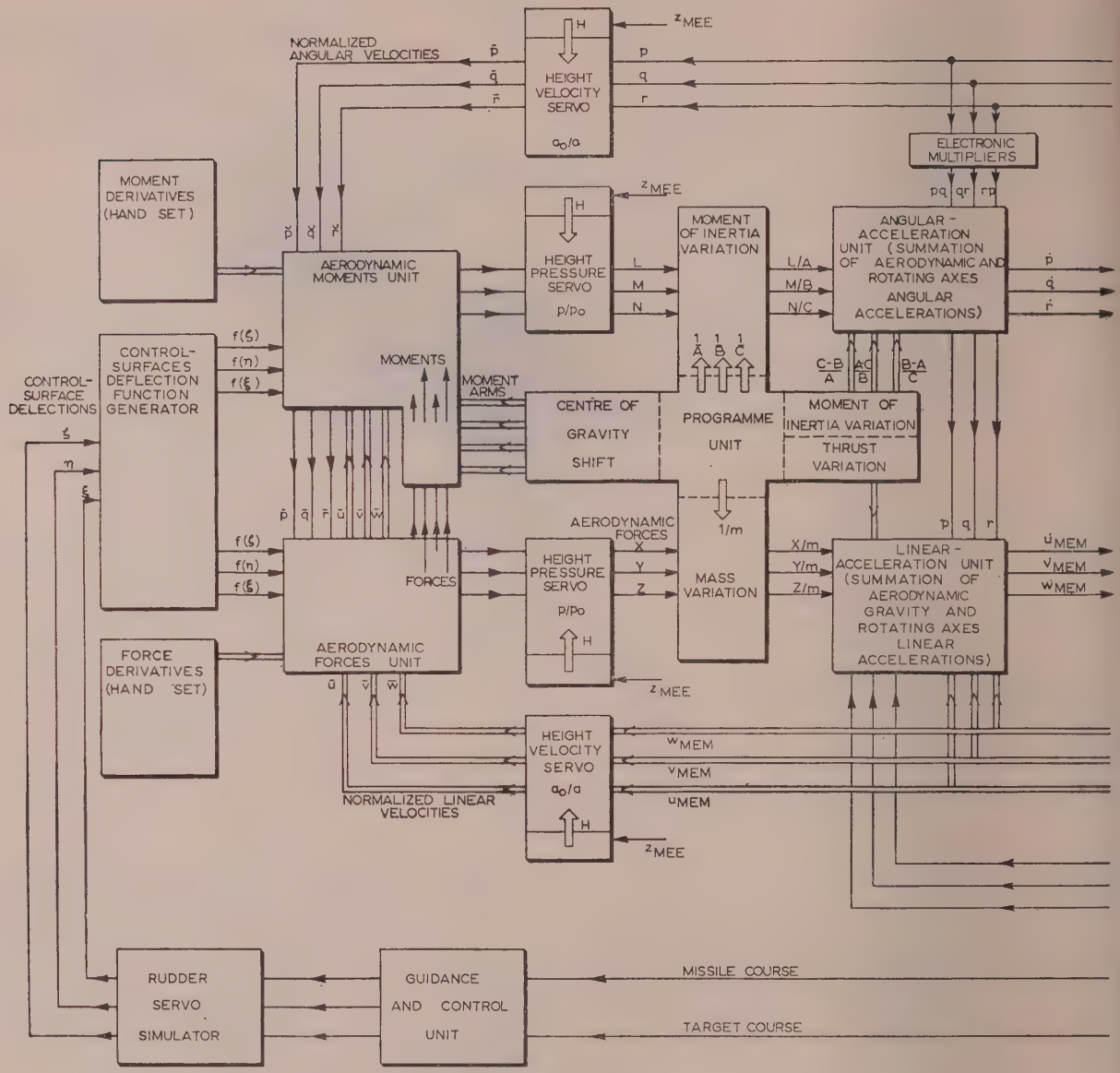


Fig. 3.—Block diagram of overall system of Tridac.

The control-surface deflections relative to the missile body are fed as functions, computed in the *control-surfaces function generator*, to the *aerodynamic-forces unit* and *aerodynamic-moments unit*, which solve the aerodynamic equations of motion of the missile in roll, pitch and yaw. The linear and angular accelerations of the missile referred to missile axes are thus computed, the variation of aerodynamic forces and moments with height above sea level being compensated by the *height-pressure servo unit*. The *programming unit*, working on a preset time basis, is used to take into account the variations with fuel consumption, etc., of mass, moments of inertia, thrust and centre-of-gravity position.

For the purpose of axis transformation the missile is considered as being mounted in a set of orthogonal gimbals, as shown in Fig. 4. Mathematically only three gimbals are required, but in order to avoid mathematical singularities (corresponding to a condition of "gimbal lock") a fourth gimbal, referred to as the "redundant gimbal" is added and orientated in such a way that this difficulty is avoided. The *synthetic-gimbal system*, which simulates the hypothetical mounting holding the missile, consists of a set of four mechanically independent shafts, each hydraulically servo-driven so that the angular positions of the shafts represent the angles between successive gimbal frames. Each synthetic-gimbal shaft carries an array of sine and cosine resolvers which enable the transformation matrices between the missile and earth axes to be built up. In particular, the *matrix unit* produces voltages proportional to the nine direction cosines which link the missile and earth-axes systems.

In the *angular-acceleration unit*, the angular accelerations, $\dot{p}$, $\dot{q}$ and $\dot{r}$, are computed with reference to non-rotating axes coincident with the instantaneous position of the missile axes (Euler's equations). This is done by summing the accelerations

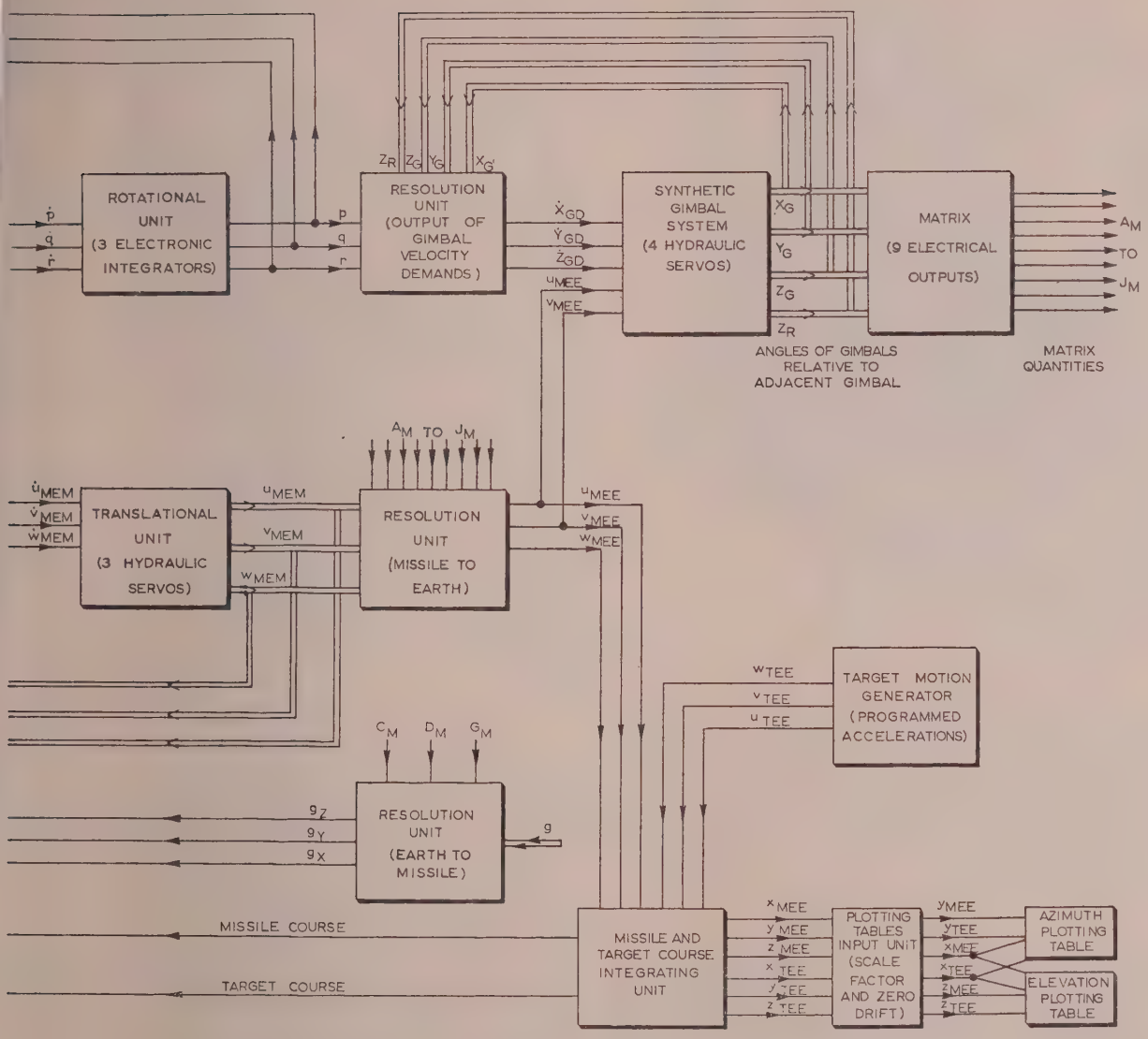


Fig. 3.—*Continued.*

due to the aerodynamic moments with those due to the angular rates of rotation of the missile axes.

In the *rotational unit* the angular accelerations are integrated electronically to angular velocities, which are converted into gimbal angular-velocity demands in a resolution unit by referring them to gimbal axes, and they are then fed to the hydraulic integrating servo mechanisms which drive the synthetic-gimbal shafts.

In the *linear-acceleration unit* the linear accelerations, $\dot{u}$, $\dot{v}$, and $\dot{w}$, are computed with reference to non-rotating axes stationary in space (Coriolis equations) by summing the accelerations due to the aerodynamic and gravity forces with those due to the missile axes being simultaneously rotated and translated.

In the *translational unit* the missile linear accelerations are integrated by hydraulic servo mechanisms to give shaft positions representing the linear velocities, u, v, and w, along the missile axes. Potentiometers mounted on these shafts are used in the aerodynamic units, where quantities dependent on u, v, and w (or on u, v, and w as modified by the local speed of sound, i.e. by a function of height, in the *height velocity servo unit*) have to be computed. For comparison with the target motion the missile velocities along the earth axes are required; these are obtained from the velocities along the missile axes using the matrix signals.

Voltages representing the programmed target component accelerations along the fixed earth axes are generated and integrated to velocities in the *target-motion generator*. The missile and target velocities along the earth axes can be integrated electronically in the *missile- and target-course integrating unit* to give displacements. The methods of using these quantities for controlling the missile motion and for setting up the *guidance and control-system simulators* vary considerably, but the result for a missile with moving control surfaces is always deflection

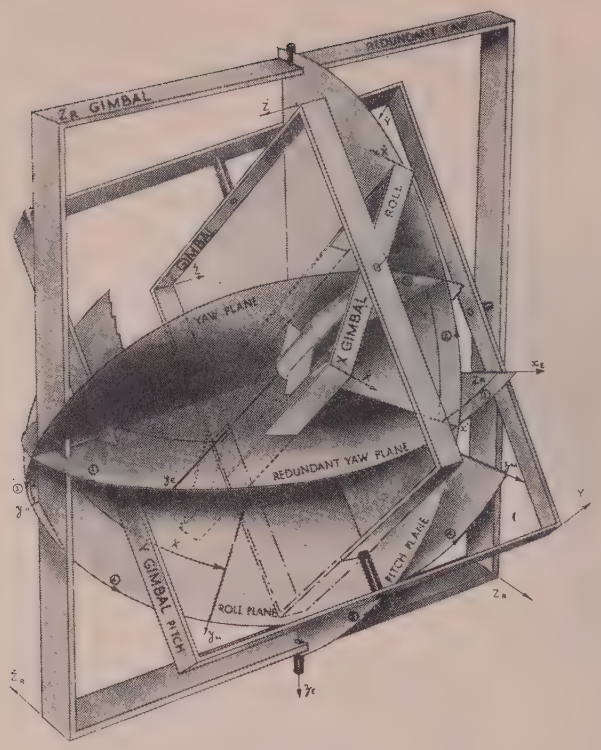


Fig. 4.—Four-axis gimbal system.

demands, which, through the actual actuators or their simulators, produce control-surface deflections which are fed to the aerodynamic units to close the outer simulator loop. There is a need for further axis transformation in this stage, since the kinematics section deals with both the missile and target as points moving relative to earth, and it is necessary to refer their relative positions to missile axes to provide control signals for the missile.

(3.3) Resolution Methods

The process of axis transformation is required to convert quantities such as displacements and velocities from one set of three axes to another, to convert missile angular velocities to gimbal velocity demands, to obtain gravity components, etc. Axis transformations are generally from earth to missile axes and vice versa, but other conversions, such as those from radar axes to missile axes, may also be required.

For 3-axis resolutions nine component quantities are necessary, namely the direction cosines, A_M to J_M, of one axis system relative to the other. Fig. 5 shows on a computing diagram the method adopted to generate, as nine electrical quantities A_M to J_M, a 3-axis transformation matrix which can be used for missile-to-earth axis transformations where the quantities to be resolved are represented by shaft positions. The four separate shafts of the synthetic-gimbal system (X_G, Y_G, Z_G, and Z_R), each driven by a high-performance hydraulic servo-motor, are fitted with a number of potentiometers which reproduce the sine or cosine of the angle of one gimbal relative to the adjacent one, as shown in Fig. 4. The generation of each term of the matrix is a process of multiplication by sine or cosine and summation, and the computing diagram can be converted to a circuit diagram for setting up the simulator by the methods of Fig. 6. A different matrix is required if a different arrangement of gimbals is needed, and removable patch panels are fitted to the cabinets so that the connections for any gimbal arrangement can be stored.

To set up as gimbal angles the four shafts of the hydraulic integrating servo mechanisms, it is necessary to derive the gimbal angular-velocity demands from the known missile angular velocities. In this process, use is made of the sine and cosine potentiometers on the gimbal shafts in making the necessary resolutions of the missile angular velocity.

The redundant gimbal shaft is driven in such a way as to prevent excessive rotation of the Z_G shaft, since this carries a secant potentiometer. The redundant gimbal need not be used, which thereby considerably simplifies the matrix expressions, if the angular travel of the missile or aircraft to be simulated is small.

(3.4) Aerodynamic Equations

The aerodynamics of aircraft and missiles vary considerably with their configuration, but examples of the general equations for a cruciform configuration are given in Appendix 11. In general, the three forces (X, Y and Z) along, and the three moments (L, M and N) about, the roll, pitch and yaw axes of the missile are computed by summing the components due to each control-surface deflection, linear velocity and angular velocity, together with any cross-coupling components.

The problem of setting up the aerodynamic section of the simulator for a particular configuration is largely that of converting the aerodynamic data into a tractable form. Such data are not generally known with great accuracy, and consequently extreme accuracy of simulation is not generally justified. Non-linear data may in most cases be adequately represented by the first few terms of a power series. In some cases simpler forms,[3] which ignore some non-linear and cross-coupling terms, can be used.

(3.5) Auto-pilots

The real auto-pilot normally accepts a guidance demand, converts it into a control-surface deflection, and measures the resulting missile motion by missile-borne instruments. The oscillations of the missile about its centre of gravity and the steady-state accelerations developed in response to given guidance demands may be controlled by feedback from missile-borne instruments, such as accelerometers, gyroscopes, etc.

In Tridac, real-motion-sensitive instruments cannot be included, and such instruments have to be simulated. If a true gimbal system carrying a platform were added, real-angular-motion measuring instruments could be subjected to angular motion, but linear-motion measuring instruments would still have to be simulated.

Facilities are, however, available for transporting a complete missile into the simulator buildings so that parts of it, such as the actuator, can be included in the simulation.

(4) COMPUTING METHODS AND COMPONENTS

(4.1) Mathematical Operations and Computing Methods

The fundamental mathematical operations carried out in Tridac are those of addition and subtraction, multiplication and division, resolution, integration with respect to time, the generation of functions of a single variable, and the generation of operational transforms with respect to time.

The principles of d.c. analogue computing are well known,[4,5,6] and the methods of computation adopted in Tridac to carry out the mathematical operations are illustrated in Fig. 6, where A denotes a very-high-gain d.c. amplifier. Typical equations, which are solved by the application of these methods, are given in Appendix 11.

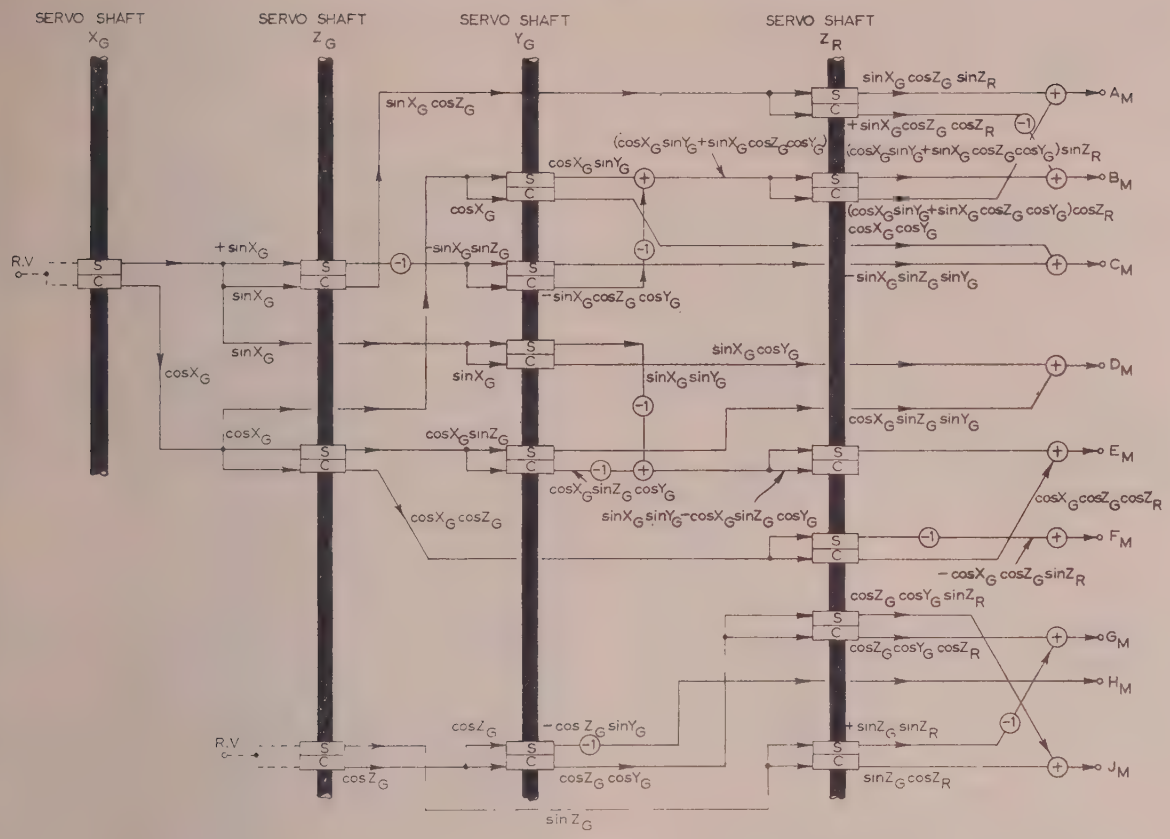


Fig. 5.—Three-axis transformation matrix generator.

$A_M = \sin X_G \cos Z_G \sin Z_R + (\cos X_G \sin Y_G + \sin X_G \sin Z_G \cos Y_G) \cos Z_R$
$B_M = -\sin X_G \cos Z_G \cos Z_R - (\cos X_G \sin Y_G + \sin X_G \sin Z_G \cos Y_G) \sin Z_R$
$C_M = \cos X_G \cos Y_G - \sin X_G \sin Z_G \sin Y_G$
$D_M = \sin X_G \cos Y_G + \cos X_G \sin Z_G \sin Y_G$
$E_M = \cos X_G \cos Z_G \cos Z_R + (\sin X_G \sin Y_G - \cos X_G \sin Z_G \cos Y_G) \sin Z_R$
$F_M = -\cos X_G \cos Z_G \sin Z_R + (\sin X_G \sin Y_G - \cos X_G \sin Z_G \cos Y_G) \cos Z_R$
$G_M = -\cos Z_G \sin Y_G$
$H_M = -\sin Z_G \sin Z_R + \cos Z_G \cos Y_G \cos Z_R$
$J_M = \sin Z_G \cos Z_R + \cos Z_G \cos Y_G \sin Z_R$

R.V. = Reference voltage.

(−1) Sign-reversal process

S / C Resolution by {sine / cosine

The problem variables are in most cases represented by voltages, and each computing unit is arranged to have an operating output range of ± 30 volts. In general, the computing units are designed to give errors which are not more than 0·1% of the full output of the unit.

New types of computing units and drift-corrected amplifiers have been developed to meet the demands for high accuracy; one problem has been to obtain in adequate quantities and at reasonable price the highly accurate components, such as resistors, which are required to maintain the overall accuracy.

(4.2) Electronic Computing Units

The basic electronic computing unit is a d.c. feedback amplifier. The accuracy requirement demands that this amplifier shall have a high gain over the frequency band from zero to several hundred cycles per second, besides very good zero stability to avoid integration errors and the necessity of constant manual resetting. To achieve this, a special amplifier is used consisting of a d.c. amplifier and a drift-correcting a.c. amplifier.[7] The feedback drift error voltage at the input of the d.c. amplifier is modulated, amplified, rectified, smoothed and fed back to the d.c. amplifier to reduce the error. The drift of the d.c. amplifier, which has a good frequency response, is thus corrected by the a.c. amplifier, which has a low drift but a poor frequency response. The computational accuracy of the complete drift-corrected amplifier is largely determined by that of the resistors in the feedback and input circuits [R_I and R_F of Fig. 6(*a*)]. These resistors are wire-wound and sealed in capsules to ensure both accuracy and stability. The allowed error in resistance is 0·1%.

The circuit of the d.c. amplifier is shown in Fig. 7. It comprises three stages of amplification and a cathode-follower output stage giving an overall internal gain of 60 000 and a very low output impedance. The input stage is a cathode-coupled pair, the first valve of which is chosen to keep the grid current below 10^{-11} amp. The drift-correcting voltage from the a.c. amplifier is applied to the grid of the second valve. By using either one or three valves in the cathode-follower output stage, maximum output currents of 4 or 12 mA are available. The utilized output range of the amplifier is ± 30 volts.

Fig. 8 is a block schematic of an amplifier with its drift-corrector unit, in which the amplifier is shown with zero input signal and with the drift potential appearing as an injected

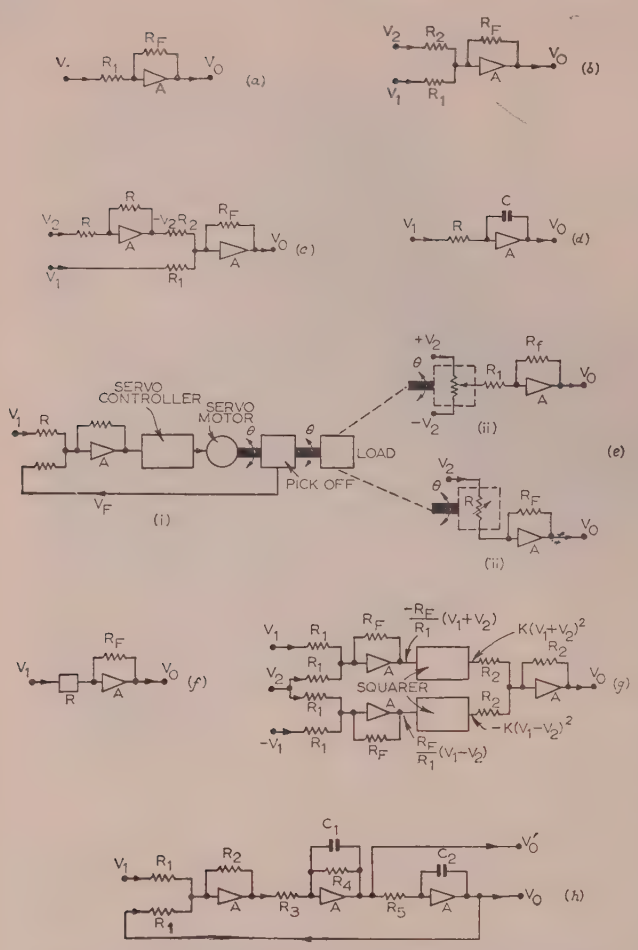


Fig. 6.—Basic analogue computing methods used in Tridac.

(*a*) Multiplication by a constant. $V_0 = -V_1R_F/R_1$.
(*b*) Addition. $V_0 = -R_F(V_1/R_1 + V_2/R_2)$.
(*c*) Subtraction. $V_0 = -R_F(V_1/R_1 - V_2/R_2)$.
(*d*) Integration. $V_0 = -1/CR\int V_1dt$.
(*e*) Shaft-setting servo system.
(i) For $V_F \propto \theta, \theta_1 \propto V_1$.
or $V_F \propto \theta_1, \theta \propto (\int V_1dt)$.
(ii) Servo load: multiplication
$V_0 \propto V_2 \times \theta \propto V_2 \times V_1$
(iii) Servo load: division
$V_0 = (-R_FV_2/R_1) \propto V_2/V_1$, since $R_1 \propto \theta \propto V_1$
(*f*) Function generator using non-linear resistances.
R = Non-linear resistance (function of V_1).
$V_0 = -V_1R_F/R = f(V_1)$, e.g. $V_0 = V_1^2$ [see Fig. 13(*c*)].
(*g*) Multiplication using squarers.
$V_0 = -K[(V_1 + V_2)^2 - (V_1 - V_2)^2] = -4KV_1V_2$.
(*h*) Simulation of a typical operational transfer function.

$$V_0 = \frac{-\omega_n^2}{p^2 + 2\zeta\omega_n p + \omega_n^2}V_1, \quad V_0' = -C_2R_5pV_0$$

$$\omega_n^2 = \frac{R_2}{R_1R_3R_5C_1C_2}, \quad 2\zeta\omega_n = \frac{1}{C_1R_4}$$

potential V_D. The drift-corrector unit has a gain β, and the d.c. amplifier a gain α.

With the corrector amplifier disconnected the drift error voltage, V_{01}, appearing at the amplifier output is

$$V_{01} = \left[\frac{(Z_1 + Z_2)\alpha}{Z_1 + Z_2(1 + \alpha)}\right]V_D$$

$$\simeq \left(\frac{Z_1 + Z_2}{Z_2}\right)V_D \text{ when } \alpha \text{ is large.}$$

When the amplifier gain is large the drift is therefore not reduced by a further increase of gain.

With the drift-corrector unit connected the drift error, V_{02}, is

$$V_{02} = \frac{1}{1 + \beta}\left(\frac{Z_1}{Z_3} + \frac{Z_1 + Z_2}{Z_2}\right)V_D$$

In practice the decrease in drift achieved by the introduction of the $(1 + \beta)$ term more than offsets the increase in drift caused by the introduction of the input impedance, Z_3, of the corrector unit. A limit to the drift correction obtainable is, however, set by the drift errors introduced by the modulation and demodulation system.

The error voltage at the input to the d.c. amplifier may be converted to alternating current either by a relay chopper unit or a magnetic modulator unit. The latter gives poorer drift correction but requires less maintenance because it has no moving parts. The relay chopper unit is used at points where the drift must be kept as small as possible, e.g. in integrators and in reproducing exact geometrical relationships, whereas the magnetic modulator unit is used at less critical points, e.g. in reproducing aerodynamic relationships.

A diagram of the relay chopper a.c. amplifier is shown in Fig. 9. The relay vibrates at 400 c/s, and modulates and demodulates the input and output respectively. The amplifier comprises two stages of amplification and a cathode-follower, and gives a low-frequency gain of about 1 000. The drift error at the output of the d.c. amplifier is held with this corrector to less than 1 mV when the overall gain is unity.

The magnetic-modulator type of corrector is shown in Fig. 10. The magnetic modulator is essentially a double iron-cored a.c.-excited unit of conventional design.[8,9] The a.c. coils are wound one on each core whilst the control winding covers both cores. The direct-voltage input is applied to the control winding, thus causing current at even harmonics of the excitation frequency to flow in that coil, the algebraic sum of the positive and negative even-harmonics being approximately proportional to the applied direct voltage. The output of the modulator appears across the resistor r. These even harmonics are then amplified in a valve amplifier and detected in a phase-sensitive peak rectifier so that the output is a direct voltage proportional to the input, and as such it may be applied to the d.c. amplifier as a drift-correcting signal. This unit maintains the drift error at the output of the d.c. amplifier within 2 mV when the overall gain is unity.

The integrator units in Tridac consist of standard d.c. amplifiers and relay drift correctors as described above, together with built-in capacitors and resistors to give the desired time-constants. Relay circuits are incorporated to set-in the initial values of the integrator output. The capacitors have a polystyrene dielectric which combines high insulation resistance with very low absorption.

Non-linear functions are, in general, generated by diode networks[10,11] since this is the only known method of obtaining the required accuracy in certain units. They are used in Tridac to simulate such effects as saturation and dead space, as shown in Figs. 11(*a*) and 11(*b*), and to generate such functions as $y = Kx^2$, $y = K/x$ and $y = K\sin x$, where both inputs and outputs are voltages. Squarer units, producing $y = Kx^2$, are used in circuits for multiplication by making use of the identity $(x + y)^2 - (x - y)^2 \equiv 4xy$. These functions are usually generated by resistor-diode combinations in the input and feedback paths of d.c. amplifiers. In this way the external gain can be made a non-linear function of the input variable voltage. Functions generated in this way are approximations to curves by straight-line segments, and the computational accuracy depends partially on the number of lines, i.e. the number of

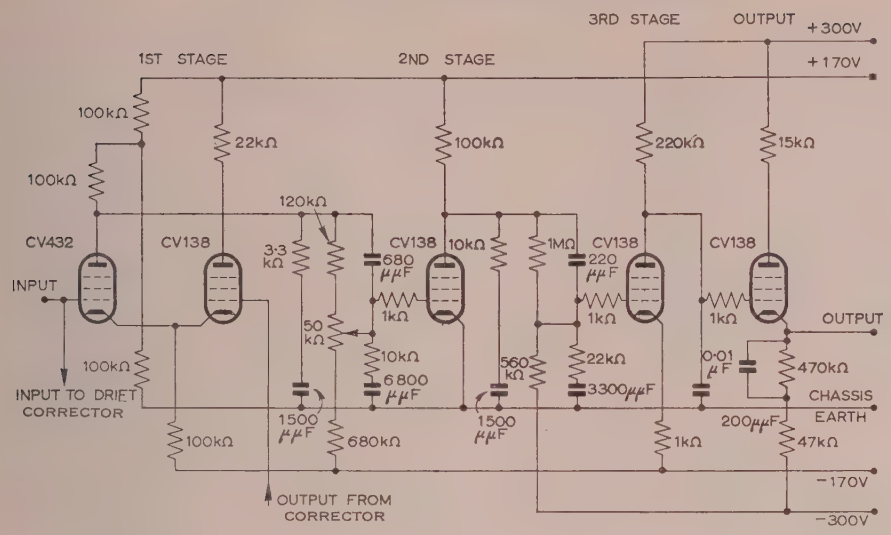


Fig. 7.—Circuit of d.c. amplifier.

diodes, used. Fig. 11(*c*) shows a typical circuit in which the gain is made to be an increasing function of the input magnitude. This type is used as a negative squarer with $V_0 = -KV_1^2$ when V_1 is negative. A positive squarer giving $V_0 = -KV_1^2$, when V_1 is positive, can be made by reversing the diode connections.

However, when errors as small as 0·1% are sought in function generation it is necessary to use well-stabilized bias voltages and d.c. heater supplies and to avoid errors due to the heater-to-cathode resistance not being infinite. Such considerations weighed against a positive squarer made by merely reversing the diodes of a negative squarer, and led instead to the use of a reversing amplifier as shown in the complete squarer circuit of Fig. 11(*d*). This circuit produces $V_0 = -KV_1^2$, for V_1 positive or negative. With an electronic multiplier using these circuits, products may be computed to an accuracy of 0·2% of full scale.

In general, the accuracy of the electronic computing units of Tridac depends almost entirely on the accuracy of resistors and capacitors used in input and feedback circuits and on potentiometers used for setting coefficients and scale factors. The summation resistors are wire-wound and grouped to avoid the effects of differential temperatures. Where linearity between shaft position and the output potential of potentiometers must be maintained, a low-impedance (7·5-kilohm) potentiometer with cam correction feeds a 2-megohm summation resistor. The required properties of the integrating capacitor have already been mentioned. It is not necessary that the absolute value of its capacitance be accurate, however, since the time-constant is fixed by adjusting the associated resistance.

(4.3) Servo Mechanisms and Mechanical Computing Units

In addition to the purely electronic computing units, servo mechanisms and associated electro-mechanical computing units are used. The advantage of using a servo mechanism is that it allows one quantity to be multiplied by a number of other quantities by using the servo output shaft to drive ganged potentiometers supplied with different voltages. The disadvantage of this method of computing is the relatively slow speed of response owing to inertia loads, and a major requirement for Tridac was a servo mechanism with a response sufficiently rapid to obviate computing on an extended time scale. When the input is only varying slowly, as with missile height and range of

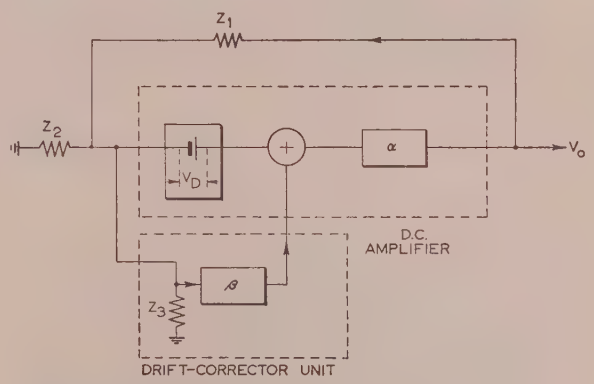


Fig. 8.—Block diagram of amplifier and drift corrector.

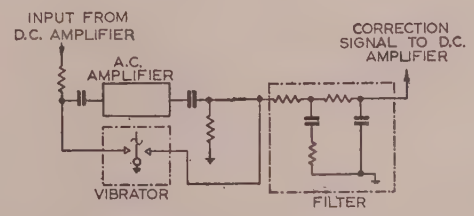


Fig. 9.—Relay drift corrector.

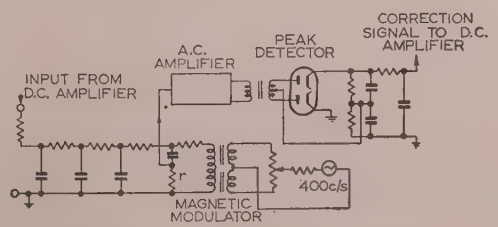


Fig. 10.—Magnetic-modulator drift corrector.

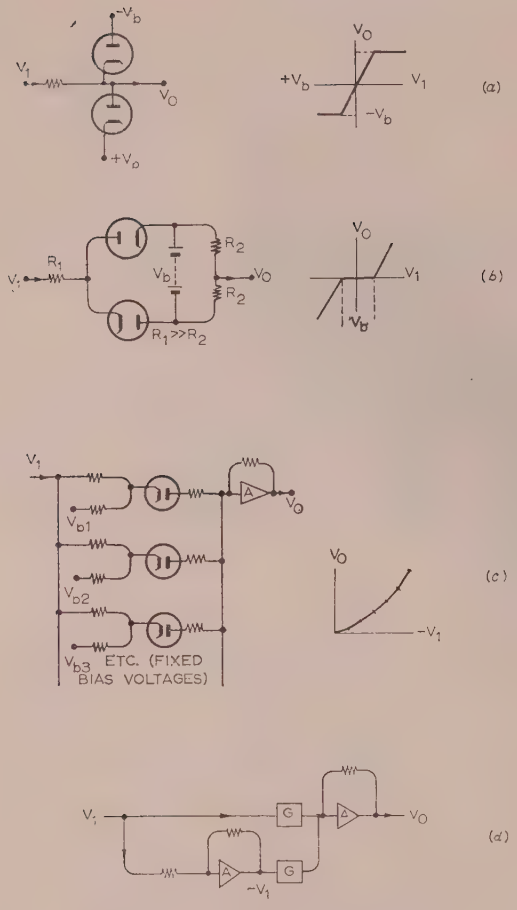


Fig. 11.—Non-linear diode units.
(*a*) Simulation of saturation.
(*b*) Simulation of dead space.
(*c*) Generation of function with increasing slope, e.g. $V_0 = KV_1^2$.
(*d*) Complete squarer. *G* is squarer-diode circuit of (*c*).

target, electrical servo mechanisms can be used, but for the rapidly varying quantities, such as the output of the servo mechanism which represents roll angle of the missile, high-performance hydraulic servo mechanisms are a necessity.

There are six electrical servo mechanisms, all of similar and conventional design. Each uses a d.c. split-field motor with potentiometer feedback. Tacho-generator output and passive networks in both forward and feedback paths provide stabilization and improve performance at low frequencies. The load is usually a bank of potentiometer wipers, geared to the output shaft either directly or through a resolving mechanism, or else a cam cut to give, for example, air pressure as a function of height. There are two such servo mechanisms representing height of missile, two representing ranges of missile and target, and two representing the angles of a ground radar dish tracking a target.

Hydraulic servo mechanisms are used where higher output rates are expected and where time lags are less tolerable. Three, which are used for simple multiplication, represent the linear velocities of the missile along its three principal axes. Four others represent the synthetic-gimbal system and are used for multiplication by trigonometrical ratios to perform axis resolution. Each of these servo mechanisms provides an overall integration between input and output. Two further servo mechanisms, also used for resolution and acting as position follow-up units, represent the gimbal angles of a homing eye mounted in the missile.

A block schematic of a hydraulic integrating servo mechanism is shown in Fig. 12, in which the input voltage, V_1, is proportional to some rate, $\dot{x}$, and the output shaft rotation is proportional to x. The hydraulic unit has two stages, one at low pressure (250 lb/in²) and one at high pressure (2000 lb/in²). The low-pressure stage consists of a torque motor driving a pilot valve, to which dither is applied to avoid stiction effects; the pilot valve controls in turn the main valve through the driving piston. The high-pressure stage consists of a vane-type hydraulic motor, controlled by the main valve, which drives a shaft geared to the load. Each stage represents, in the steady-state condition, an integration with a quadratic lag. A subsidiary loop is formed by feeding back, from a pick-off unit, the main valve displacement. Feedback for the main loop is derived from a tacho-generator on the output shaft of the vane motor; to obtain the required accuracy of integration the tacho-generators require to be specially selected so that the departure from linearity does not exceed 0·1%.

An input voltage to the subsidiary loop gives rise, in the steady state, to a displacement of the main valve, and consequently a rate to the motor. For the servo mechanisms to have zero steady-state velocity lag at the output shaft, it is necessary to have integration in the main loop; this is provided by capacitance feedback on the error-detecting amplifier. By using filters in the feedback and error channels the main-loop natural frequency is made to be about 40 c/s. At 5 c/s the output/input error does not exceed 5° in phase and 5% in amplitude, and in response to a step corresponding to 180° shaft rotation, the first overshoot does not exceed 20% and the time taken to settle down is 0·12 sec. The maximum angular speed of the output shaft is 100 rad/sec and the maximum acceleration about 2 500 rad/sec².

Since this servo mechanism is an integrator, it is necessary to set up the initial position of the output shaft. This is achieved by comparing an initial voltage signal with a voltage fed back from a reset potentiometer on the shaft, and using the difference error signal as an input. Switching from initial conditions is by relays, which are automatically operated at the start of a run.

The hydraulic servo mechanism may be used to drive a sine-cosine potentiometer unit. Each unit consists of a central shaft carrying a swashplate (see Fig. 13) which drives two sets of 12 wipers, each with a travel of ±2″, on linear potentiometer elements. If a constant voltage is applied across all the potentiometer elements and the angular rotation of the main shaft is θ, the voltages on the two sets of wipers are respectively proportional to $\sin\theta$ and $\cos\theta$. The maximum error in a sine-cosine unit is about 0·2%.

(4.4) Programming Unit for Computer

The input to the machine when considered as a closed-loop system, as shown in Fig. 2, is the target motion. Both the target motion and certain other problem variables, such as the motor thrust and the changes in mass, moments of inertia, and position of centre of gravity due to fuel consumption, will usually be known functions of time. When operating on a real time scale these quantities, which form the only variable inputs to the machine during a simulated flight, require to be varied at their real rate in order to programme the computer. If it is desired to operate on an extended time scale, it is only necessary to decrease the rate of change of these inputs and reset the simulator time-constants.

An additional input to the machine is the injection of noise such as would exist, for example, in a radar system. Facilities

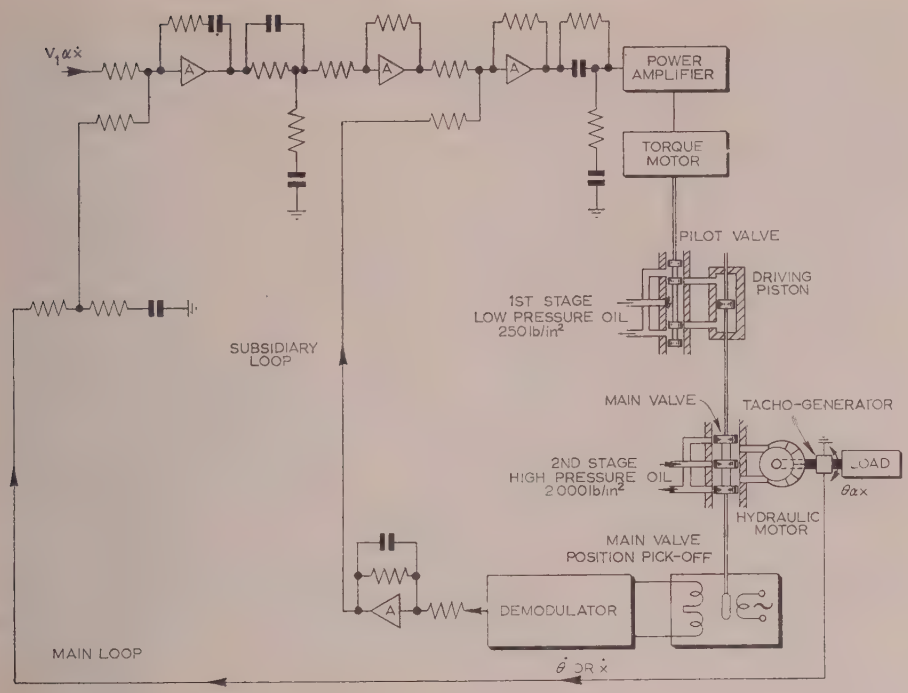


Fig. 12.—Block diagram of hydraulic servo mechanism.

are available for injecting such noise at suitable points. The noise signals for the simulator may be records of actual noise, played back from photographic film or magnetic tape, or they may be generated by a synthetic-noise machine consisting of a source of "white noise" (i.e. noise of uniform spectral density in the relevant frequency band) followed by suitable filters.

A function of time can be approximated by using a uniselector operated from time pulses. For example, the motion of the moving target is generated by step approximation to the three components of acceleration, followed by double integration to obtain the position components as smooth, continuous voltages.

In addition to generating functions during a problem run, the programming unit is used to set up the initial outputs of electronic and servo integrators, to reset and start recorders and also to reset the zeros and monitor the drift in the hydraulic servo mechanisms. This is done, during a 25 sec period before the problem is started, by relay circuits operated from a uniselector.

Initial integrator outputs are set on potentiometers, the electronic integrators being temporarily converted to unity-gain amplifiers, and the hydraulic integrators being temporarily converted to position servo mechanisms, until the problem begins. With the synthetic-gimbal shafts it is necessary at the start of the problem for the shafts to be not only in their correct initial positions but also moving at their correct angular rates. This is achieved by starting the integrating servo mechanisms from a position setting at $t = -1$ sec with specially computed input signals, so that at $t = 0$ the shafts are in the correct positions and moving with the correct rates. At $t = 0$ the normal servo-mechanism inputs are applied simultaneously with the starting of the remainder of the simulation.

Each of the hydraulic servo mechanisms is fitted with a reset unit which applies a drift-correcting voltage when the input is earthed. The programming unit switches in this unit and then, when the servo mechanism is switched to a position setting, it monitors the error signal before switching the mechanism to integrate.

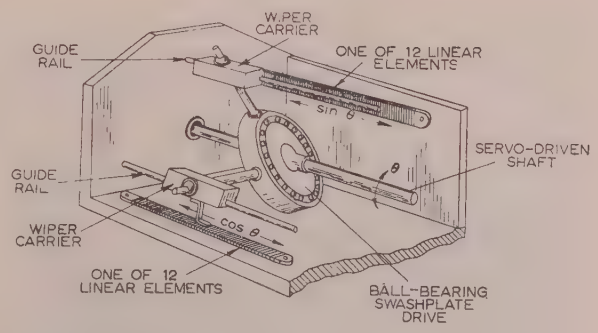


Fig. 13.—Swashplate mechanism.

The programme may not be started unless all of the many interlock circuits are in their correct state, and it is automatically stopped at the end of the running period or when a fault occurs, or, in some cases, when a particular variable passes through a minimum. On such occasions the machine is reset by a manually-operated switch. The problem may be repeated from the beginning by another control.

(4.5) Recording and Displaying of Results

Multi-channel recorders, plotting tables and special-purpose recorders are fitted to record the results of trials for quantitative analysis, and various displays are fitted for qualitative analysis. The displays are intended to decrease the amount of quantitative

analysis required, and the relevant variables are recorded on a multi-channel tape recorder so that all or part of any run can be restudied.

Four 3-channel variable-paper-speed moving-coil recorders, with a flat response at an amplitude of $\pm 1 \cdot 5$ in up to 15 c/s and recording either by ink or h.v. spark on special paper, are built into the machine; the selection of the quantities to be recorded is by plug connections on the control desk. For quantities which vary more rapidly, other portable recorders, one of the moving-coil type with four channels capable of recording by h.v. spark at an amplitude of $\pm 1 \cdot 5$ cm up to 90 c/s, and another of the Duddell oscillograph type with three channels capable of recording photographically up to higher frequencies, are available.

Two plotting tables, each fitted with two pens driven by electrical servo mechanisms, record the plan and elevation positions of target and missile on stationary charts, each of which measures 82 in × 32in. Special-purpose units record the r.m.s. value of a variable over a period and the maximum or minimum value of a variable. Two of each of these units, which may be used for measuring miss distance, are provided.

The simulated flight may be observed by means of a number of cathode-ray-tube displays, which are mounted on the control desk and show plan and elevation views of missile and target. There is also a mobile 3-dimensional display, in which alternate right-eye and left-eye views are projected from a cathode-ray tube, through a synchronized spinning disc which polarizes the views, on to a screen. The observer views this screen through polarized spectacles which select the correct image for each eye in the conventional manner to give the illusion of a third dimension.

(5) INSTALLATION

(5.1) General Layout

To facilitate construction and maintenance, the majority of the electronic units (amplifiers, stabilizers, correctors, etc.) are

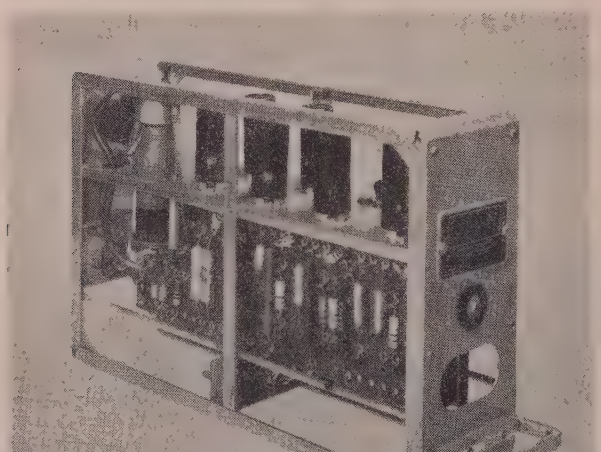


Fig. 14.—Standard electronic unit.

housed in standard chassis units (12 in × 8 in × 3 in, see Fig. 14). These "brick units," of which there are some 2000, slide into colour-coded trays to engage power and signal connections through separate plugs and sockets. The trays are mounted at the back of cabinets 27 in wide, on the front of which are panels of coefficient potentiometers, patchboards and monitoring equipment. Cabinets are assembled in sets of four (three generally contain computing units and one contains stabilizers) on a steel structure to form a "raft." There are 11 rafts in Tridac, each of which is a self-contained computer, usually of 224 units, complete with termination board, air ducts, power busbars and monitoring equipment.

The general layout of Tridac is shown in Fig. 1, and Fig. 15 shows part of the control room in which are seven rafts (in two tiers), the control desk and the plotting tables.

(5.2) Power Supplies and Protection

Tridac consumes a peak power of 600 kW, of which 200 kW is dissipated in the electronic units and 400 kW in the hydraulic systems. Motor-generator sets are used to provide various a.c. and d.c. supplies, including the unstabilized h.t. supply. The frequencies of the a.c. supplies are checked against a crystal-controlled frequency divider. The h.t. supplies are locally stabilized in each raft. Valves in stabilizers, etc., are heated by a 50 c/s a.c. supply, but valves in the computing amplifiers are heated by a d.c. supply, derived by 3-phase rectification from a 400 c/s a.c. supply and voltage-stabilized by variable transformers, which are servo-controlled by voltages fed back from the rafts.

The use of fuses for protection of individual units presents difficulties, especially in the case of stabilized h.t. and valve-heater supplies, where the source impedance would be materially increased by the fuse resistance. The fusing protection would also be uncertain, since many of the currents involved even under short-circuit conditions are so small that the appropriate fuses would not be reliable. Power protection in Tridac is accordingly limited to preventing damage to raft and cabinet cable-forms, thus avoiding the extreme complexity and consequent fault liability that would occur if full protection were applied unit by unit.

To reduce the possibility of damage by fire, the installation is protected by an extinguishing system which floods the cabinets with carbon dioxide through the air-cooling ducts. Smoke detectors are fitted in the air ducts to give aural and visual warnings, and they may also be used for automatic control of the extinguishing system.

Control contactors in the main supply lines to each raft are operated by a series of interlocked relay circuits. H.T. supplies and outputs from stabilizer units in the rafts are checked in sequence (by uniselector switching) against a standard 30-volt reference supply, an error signal closing a relay to indicate a power fault. Valve-heater supplies of several hundred amperes at 6·3 volts d.c. are monitored by d.c. transformers (magnetic amplifiers) which operate relays in the interlock circuits of the power-supply lines if an out-of-balance of ampere-turns appears between heater current and a reference bias current (Fig. 16).

The nine hydraulic servo mechanisms, shown in Fig. 17, have oil supplies delivered at two pressures, 2000 and 250 lb/in^2, by gear pumps. A single pump delivers the low-pressure oil to all of the nine units, while a separate high-pressure pump is used for each. Pumps are used in preference to an accumulator system since the duty cycle of the complete installation cannot be predicted for all applications, and also because of the possibly poor transient performance if an accumulator were used.

All the oil is continuously circulated by other pumps through forced-draught radiators for cooling and through filters for cleaning. Any pressure failures or excessive oil temperatures are indicated on the control desk, and safety interlocks are fitted.

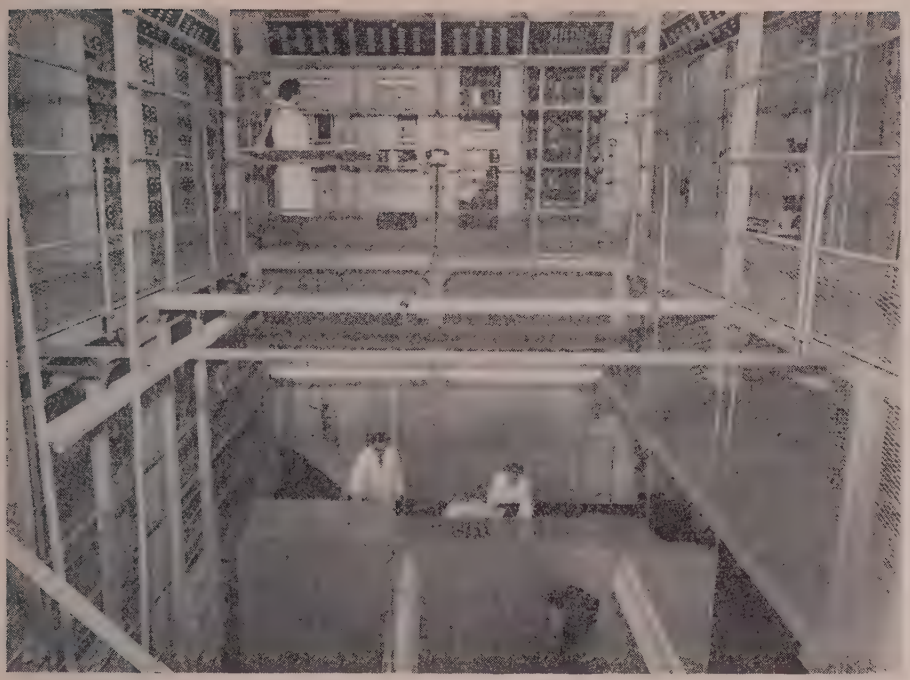


Fig. 15.—Control room.

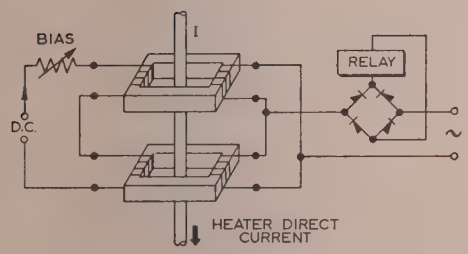


Fig. 16.—Heater-current protection.

(6) SETTING-UP OF PROBLEMS

(6.1) Interconnection of Units

Tridac has been designed and built for the special purpose of simulating 3-dimensional flight problems, and in consequence the general form of the computations is fixed. The signal connections between the computing blocks shown in the block schematic (Fig. 4) form permanent signal wiring which runs directly from one raft termination panel to another. The flexibility of the machine lies in the manner of interconnecting units within the computing sections. Three methods are used. The first, which is used where the form of the equations is expected to be changed frequently, consists of plug-and-socket connections on panels fitted to individual cabinets. The second, which is used where changes of form are fairly standard (e.g. for changes in the method of axis resolution) consists of removable patch panels which may be readily stored and quickly interchanged. Finally, detachable panels are available on which various special-purpose circuits, such as occur in control system, can be built.

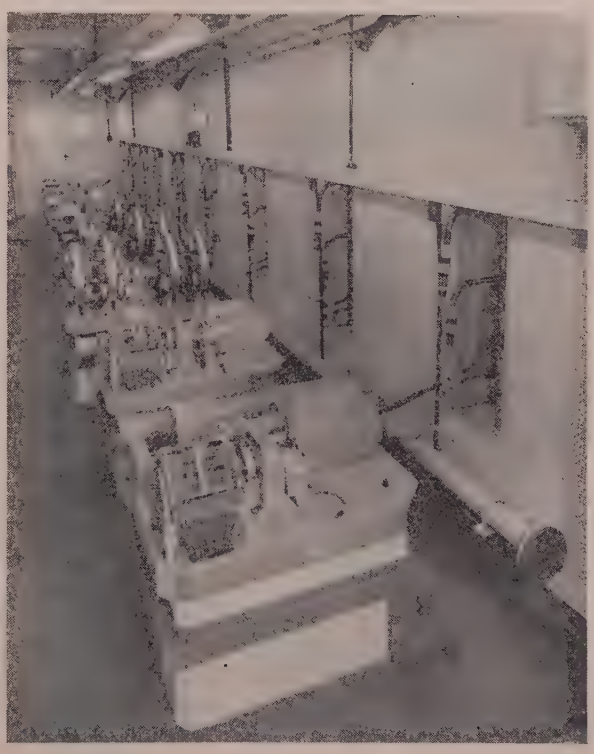


Fig. 17.—Synthetic-gimbal room.

The computational-accuracy requirements, and the size of the installation with its consequent long runs of signals cabling, make the resistance of the wires used for interconnecting rafts a significant factor. To avoid errors due to voltage drop in these wires, two measures have been taken: the first is the inclusion of trimmers on all coefficient and servo potentiometers, and the second is the use of two earth lines, a "signals earth" and a "chassis earth," joined at one central point. The signals earth carries very little current since all the h.t. currents are returned by the chassis earth line, and the resulting current here is minimized by feeding opposing current from a separate d.c. source in each raft.

(6.2) Setting-up Procedure

Each part of Tridac has an analogy in the system it is simulating, and the machine is thus most appropriately set up in a sequence corresponding to the logical build-up of the problem. At each stage in the setting-up procedure the functioning of the units involved is checked and the accuracy of simulation is assessed. Thus the flight of an aircraft or missile under control, for example, may be set up on the simulator by first setting only the units which simulate the roll motion. This is achieved by suitably interconnecting the voltages representing the quantities contributing to the rolling moment and introducing the appropriate aerodynamic coefficients. The resulting angular acceleration in roll may then be integrated to give roll rate and roll position, and the operation of this portion of the simulator can be checked against theoretical solutions for steady-state values, response to step inputs, etc., as desired.

After treating the pitch and yaw motions similarly, and checking the operation, the cross-coupling terms between the three modes of motion may be introduced one by one and the senses, etc., can be checked. Normally, the system is first set up on the basis of constant coefficients to facilitate checking, and then extended to cases of varying coefficients as required.

The motion of the aircraft or missile with respect to its own translating and rotating frame of reference is now known. To obtain the motion in a frame of reference fixed in space, the axis-resolution equipment is required; this is set up and tested as a self-contained unit and then applied to the problem of resolving the missile motion into the space frame of reference.

The problems for simulation may now be built up and checked step by step around this central core of the basic flight simulation.

Fig. 18(*a*) shows the computing diagram for the yaw-plane

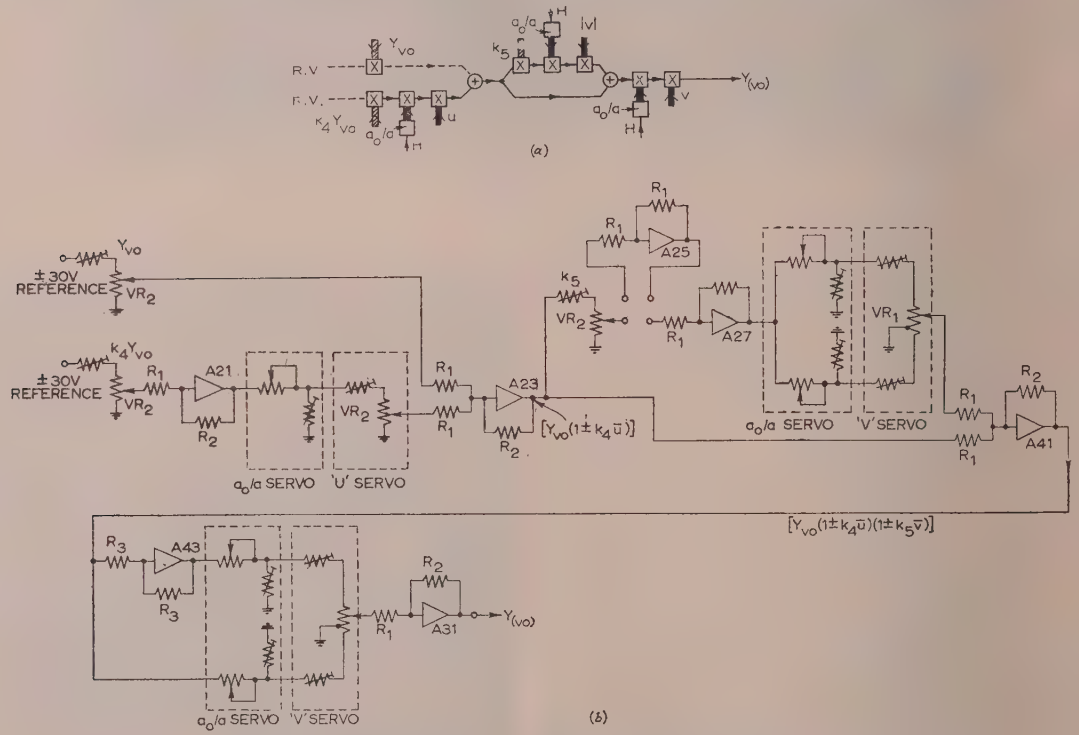


Fig. 18.—Computing and circuit diagram for yaw-plane force, Y_{v0}.

Varying mechanical signal (rotating driver shaft).
Fixed mechanical signal (hand-set shaft).
Varying electrical signal (varying voltage).
Fixed electrical signal (reference voltage).
× Multiplication process.
+ Summation process.
(−1) Sign-reversal process.
S/C Resolution by {sine, cosine}

(*a*) Computing diagram $Y_{(v0)} = Y_{v0}(1 + k_4\bar{u})\bar{v}(1 + k_5|\bar{v}|)$.
(*b*) Circuit diagram.
$R_1 = 2\,M\Omega$.
$R_2 = 2{\cdot}222\,M\Omega$.
$R_3 = 100\,k\Omega$.
$VR_1 = 15\,k\Omega$.
$VR_2 = 7{\cdot}5\,k\Omega$.
—▷— High-gain d.c. amplifier.

force, $Y_{(v0)}$, and Fig. 18(*b*) shows the corresponding circuit diagram, which is typical of the circuits used throughout Tridac. In detail, the setting-up procedure ultimately reduces to the insertion of known resistors, the selection of patch-panel connections, and the setting of coefficient and initial-condition potentiometers, etc. A scale factor for each quantity is assessed so as to give maximum swing of all variable quantities without overloading; these scale factors may require some adjustment during preliminary trials, if overloading is indicated on the monitoring system.

(7) MAINTENANCE AND CHECKING PROCEDURE

Computations carried out will be unreliable and may be misleading unless the entire simulator is known to be functioning correctly during problem runs. The rapid location of faults is therefore essential, and this, because of the size and complexity of the machine, presents many unusual problems.

A comprehensive monitoring system is used to detect and display certain fault conditions throughout the simulator while it is running. In addition to the h.t.-supply, heater-supply and hydraulic protection arrangements already mentioned, every d.c. amplifier is checked for drift in an automatic sequence prior to each run, and any drift in excess of 3 mV is registered as a fault. The output of each amplifier is also monitored so that overload conditions, corresponding to outputs in excess of 30 volts, can be detected and registered as faults.

Overloading may occur even with correctly functioning units if the scale factors for the analogue voltages are unduly large, and as any overload, however short in duration, might lead to erroneous results, the overload monitoring is operated continuously.

Where possible, automatic checks are made on the correctness of simulation. For example, the direction cosines obtained from the axis resolution unit must satisfy certain orthogonal relationships of the type $l^2 + m^2 + n^2 = 1$, and auxiliary squarers and adders may be used to monitor the unit continuously and ensure that its operation is consistent with the implied orthogonal relationships.

Sufficient spares of all electronic units are carried to enable units to be serviced regularly and faulty units to be replaced immediately. Statistical records are kept of the history of each unit so as to build up data which will indicate what faults are most likely and with what frequency they may occur.

Before use, all valves undergo an ageing process in equipment installed for the purpose. The process includes a switching cycle of heater and h.t. voltages which causes failure of any weakly welded electrodes.

(8) CONCLUDING COMMENTS

The cost of developing and manufacturing Tridac has been in the neighbourhood of £750 000, but the production costs of a similar machine would probably be about £250 000. It is clear that full use must be made of any such machine to justify the expenditure on it, but its cost has to be considered relative to the saving in full-scale trials and to the value of the technical advances it alone makes possible. For the solution of many problems a relatively simple machine, the cost of which is quite small, is adequate and the use of a large machine which could otherwise be employed on a major problem is quite uneconomic.

In the particular case of guided weapons the cost of firing expendable missiles for test purposes is so high that the resulting saving in the cost of missiles alone and the man-power to fire them can easily offset the manufacturing costs of Tridac. The ability to compute on a real time scale means that the cost of, and time required for, each simulated run will be low; this is of great importance in making a series of runs, such as may be necessary for the statistical testing of systems in the presence of such noise as may disturb the real system. Although the simulator can never fully replace flight trials it can replace some of the steps which occur during system development, thereby decreasing the number of rounds to be fired. The fact that the machine provides a working model of the system under study, and not merely the end result of a mathematical problem, facilitates the understanding of the functioning of the system.

A possible future addition to Tridac is a high-performance flight table on which auto-pilot and radar-equipment components, which are sensitive to angular motion, could be mounted, thereby further decreasing the number of development rounds to be fired. The development of such a table would, however, be a difficult and expensive engineering undertaking. Improved techniques for high-speed multiplication and for accurate curve following in one or two dimensions are also desirable.

For certain problems a higher order of accuracy than that obtainable with Tridac is desirable, and with an extensive problem it is difficult to assess the overall limit of accuracy owing to component imperfections. On the electro-mechanical side the accuracy is usually limited by the dynamic response characteristics, while on the electronic side it is in general limited by the absolute accuracy and stability of components, and, in many cases, this limit appears to have been approached.

Digital computers, the accuracies of which are not critically dependent on component values, suffer from the disadvantage of being unable to compute sufficiently rapidly to operate on a real time scale for aircraft and missile flight problems.[12] If the computing rate of digital machines could be increased, and if fast and accurate analogue-digital convertors were available, it would become possible to construct real-time-scale simulators of a higher order of accuracy than the existing all-analogue machines.

There thus appears to be plenty of scope for computing-machine designers and manufacturers in improving computing techniques, accuracy and speed. For analogue machines the limit of accuracy appears to have been approached in certain cases, unless manufacturers can produce components of higher accuracy and stability at a reasonable cost. For many applications, however, extreme accuracy is not essential, and the analogue-machine operator has much work to do in utilizing the machinery now available to him. It is felt that, apart from the guided-missile application described in the paper, there are many other applications for analogue computers in the fields of, for example, aircraft flight, human control of machines, economics and military tactics.

(9) ACKNOWLEDGMENTS

Acknowledgments are made to the Ministry of Supply, to the Controller, H.M. Stationery Office, and to Messrs. Elliott Brothers (London) Ltd. for permission to publish this paper.

Acknowledgments are also made to the many who have been associated with the work of developing, manufacturing and installing Tridac, and in particular to Messrs. C. A. A. Wass, K. C. Garner and P. M. Lavelle of the Royal Aircraft Establishment, to Mr. W. R. Thomas, formerly of the Royal Aircraft Establishment, to Mr. J. F. Coales, and Mr. W. H. Pearse, formerly of Elliott Brothers, Ltd., and Messrs. A. D. Jeffery, A. J. Wakefield, J. C. Nutter and A. Leech of Elliott Brothers, Ltd.

(10) REFERENCES

(1) HALL, A. C.: "A Generalized Analogue Computer for Flight Simulation," *Transactions of the American I.E.E.*, 1950, **69**, p. 308.

(2) GARDNER, G. W. H.: "Guided Missiles," *Chartered Mechanical Engineer*, 1955, **2**, p. 5.
(3) SPEARMAN, F. R. J.: "Aerodynamic Transfer Functions," *Journal of the Royal Aeronautical Society*, 1955, **59**, p. 743.
(4) KORN, G. A., and KORN, T. M.: "Electronic Analog Computers" (McGraw-Hill, 1952).
(5) WASS, C. A. A.: "Introduction to Electronic Analogue Computers" (Pergamon Press, London, 1955).
(6) WILLIAMS, F. C., and RITSON, F. J. V.: "Electronic Servo Simulators," *Journal I.E.E.*, 1947, **94**, Part IIA, p. 112.
(7) GOLDBERG, E. A.: "Stabilization of D.C. Amplifiers," *R.C.A. Review*, 1950, **2**, p. 296.
(8) FROST-SMITH, E. H.: "An Experimental Study of the Magnetic Amplifier," *Journal of the British Institution of Radio Engineers*, 1949, **9**, p. 358.
(9) GALE, H. M., and ATKINSON, P. D.: "A Theoretical and Experimental Study of the Series-Connected Magnetic Amplifier," *Proceedings I.E.E.*, Paper No. 778 M, November, 1948 (**96**, Part II, p. 339).
(10) BURT, E. G. C., and LANGE, O. H.: "Function Generators based on Linear Interpolation with applications to Analogue Computing," *Proceedings I.E.E.*, Monograph No. 137 M, June, 1955 (**103** C, p. 51).
(11) HORDER, C.: "New Techniques on the Anacom Electric Analogue Computer," *Transactions of the American I.E.E.*, 1950, **69**, p. 547.
(12) DIPROSE, K. V.: "Analogue Computing in Aeronautics," *Journal of the Royal Aeronautical Society*, 1955, **59**, p. 479.

(11) APPENDIX

(11.1) General Equations solved in Tridac

(11.1.1) General and Notation.

The general equations solved in Tridac are given here and in the diagrams; they comprise the aerodynamic equations, the equations of mechanics, and the gimbal system and axis-resolution equations. The guidance and control-system equations are particular to the systems under test, and no general equations can be given.

The notation is as follows:

(*a*) The standard notation, as given in textbooks on aerodynamics and Reference 3, is used for aerodynamic quantities with control-surface deflections denoted by ξ, η, ζ.

(*b*) The dimensional aerodynamic derivatives are denoted by symbols such as Y_v and L_ξ with particular values which are constants denoted by symbols such as Y_{v0} and $L_{\xi 0}$.

(*c*) The ratios of air pressure and sound velocity at height and sea level are denoted by p/p_0 and a/a_0, respectively.

(*d*) The mass of the missile is denoted by m, and the moments of inertia about the three axes by A, B and C.

(*e*) The true missile linear (u, v, w) and angular (p, q, r) velocities are normalized by multiplying by a_0/a to give $\bar{u}$, $\bar{v}$, $\bar{w}$, and $\bar{p}$, $\bar{q}$, $\bar{r}$.

(*f*) The gimbal angles are denoted by X_G, Y_G, Z_G and Z_R as shown on Fig. 4, and the matrix quantities by A_M to J_M.

(*g*) The three suffixes to any quantity denote, in order, the item associated with the quantity, the reference from which the quantity is measured, and the axis system in which the measurement is expressed.

(11.2) Equations

One of each series of equations is given below as examples of the types of equations solved.

(11.2.1) Aerodynamic Equations.

The force equation for the yaw-axis force, Y, with similar equations for the roll-axis force, X, and the pitch axis force, Z, is

$$Y = \frac{p}{p_0}\left[Y_{\xi 0}\frac{\sin 2\zeta}{2}(1 + k_3\bar{u}) + Y_{v0}(1 + k_4\bar{u})\bar{v}(1 + k_5|\bar{v}|) + Y_{v0}\bar{v}(1 + k_6\bar{u})\right]$$

The moment equation for the roll-axis moment, L, with similar equations for the pitch axis moment, M, and the yaw-axis moment, N, is

$$L = \frac{p}{p_0}\left\{ L_{00}(1 + k_{11}\bar{u}) + L_{\xi 0}\xi[(1 + k_{12}\bar{u}) - (1 + k_{13}\bar{u})(\bar{v}^2 + \bar{w}^2)] + L_{p0}\bar{p}(1 + k_{14}\bar{u}) + L_{vw0}\bar{v}\bar{w}(1 + k_{15}\bar{u})(\bar{v}^2 - \bar{w}^2)\right\}$$

(11.2.2) Equations of Mechanics

The angular-acceleration equation for $\dot{p}$ (Euler's equation), with similar equations for $\dot{q}$ and $\dot{r}$, is

$$\dot{p} = \frac{L}{A} - \frac{C - B}{A}rq$$

The linear-acceleration equation for $\dot{v}_{MEM}$ (Coriolis equation), with similar equations for $\dot{u}_{MEM}$ and $\dot{w}_{MEM}$, is

$$\dot{v}_{MEM} = \frac{Y}{m} + g_y + pw - ru$$

(11.2.3) Four-Axis Gimbal System and Axis-Resolution Equations

The equations vary with gimbal order; the values given below are for the gimbal order from earth inwards through redundant yaw Z_R, pitch Y_G, yaw Z_G, and roll X_G.

The equation for A_M, with similar equations for B_M to J_M, as given in Fig. 5, is

$$A_M = \sin X_G \cos Z_G \sin Z_R + (\cos X_G \sin Y_G + \sin X_G \sin Z_G \cos Y_G)\cos Z_R$$

The equation for the gimbal velocity demand $\dot{X}_{GD}$, with similar equations for $\dot{Y}_{GD}$ and $\dot{Z}_{GD}$, is

$$\dot{X}_{GD} = p - \dot{Y}_G \sin Z_G + \dot{Z}_R \sin Y_G \cos Z_G$$

The missile-to-earth axes resolution equations and vice versa are of the form

$$u_{MEE} = H_M u_{MEM} + F_M v_{MEM} + A_M w_{MEM}, \text{ etc.}$$

as given by the matrix equation

$$\begin{bmatrix} u_{MEE} \\ v_{MEE} \\ w_{MEE} \end{bmatrix} = \begin{bmatrix} H_M & F_M & A_M \\ J_M & E_M & B_M \\ G_M & D_M & C_M \end{bmatrix} \begin{bmatrix} u_{MEM} \\ v_{MEM} \\ w_{MEM} \end{bmatrix}$$

The gravity resolutions are of the form

$$g_X = G_M g, \text{ etc.}$$

DISCUSSION BEFORE THE INSTITUTION, 1ST DECEMBER, 1955

Mr. J. F. Coales: The paper raises important questions about the relative merits of analogue and digital computers. When I first became involved in the design of this computer, in 1948, I thought that analogue computers were on their way out and digital computers would replace them. I now think quite differently, partly because of the experience of Tridac, but partly because of other considerations met in solving non-linear problems, where one great value of the analogue computer is that a very much better appreciation is obtained of how the process being investigated really behaves. With an analogue

computer one can see what is going on in different parts of the system while it is actually happening, whereas with the digital computer, although one can make provision for doing this by printing out different results from within the process, one cannot easily see what is happening while the process is being worked out.

However, that does not necessarily mean that the decision was right in this instance. I do not wish to question that decision now, but only the reasons why we were all convinced in 1948 that Tridac must be an analogue computer.

The essential reason was that calculations based on experience with a small digital machine showed that the problems for which Tridac was required could not be solved in real time with a digital computer. The limiting factor was the time taken for axis conversion, which involves rather complicated simultaneous equations, involving a large number of multiplications of sines and cosines. The small digital computer, which was available, was doing this at an adequate speed for conventional gunnery but not in anything like the real time for this problem. The digital computer was quite small compared with machines that are now available. However, it contained 700 valves, and it would have meant constructing a machine many times larger to do this problem in real time. It was thought that a digital machine was just not practicable; a serial machine would be too slow and a parallel machine too large.

In 1948, we did not consider carrying out the axis conversion by deriving the direction cosines by use of the angular velocity of the moving frame of reference to obtain the derivatives of the direction cosines, and then obtaining the direction cosines themselves by simple integration. There is then no problem of gimbal lock, because the rates used in the calculation never become infinite, and the number of multiplications to be made and the amount of information to be stored is very much reduced. If this method had been considered, the required digital computer might have appeared to be much more reasonable.

A parallel digital computing machine costing £750 000, with a peak power of 600 kW and the equivalent mean power of Tridac, would be very large and could undoubtedly carry out the calculations required, and many more besides, but in 1948 the estimated cost of this equipment was not a tenth of £750 000.

I do not wish to suggest that such a machine would be a suitable replacement for Tridac, since there are other reasons for wanting an analogue machine, but the comparison is worth considering because there is little doubt that a digital machine would be easier to get into operation and maintain than an analogue machine. Tridac has about 2 000 amplifiers and some 3 000 valves, so that the equivalent digital machine would be unlikely to have more circuit elements, and would have the important advantage that their tolerances and stability would not be so critical.

The great problem in analogue computing is still that of maintaining the drift of the d.c. amplifiers and associated circuits at a sufficiently low level. Although this seems to have been satisfactorily solved in Tridac, it may still prove an embarrassment in maintaining the machine.

On the question of axis conversion and what we call 'flight tables', there is a challenge to the mechanical engineers. Most of these large simulators for guided weapons have had, as a requirement, a flight table, i.e. a moving platform whose movements are those of the missile, on which component parts of a guidance system can be placed and tested in the system. This was one of the reasons put forward for requiring the problem to be done in real time. Unfortunately when we try to design a set of gimbals to carry an instrument, and if the instrument is much larger than a lady's wrist watch, we find that, because there have to be four gimbals each one larger than the next, the forces on the outer gimbal rings are such that we cannot get adequate rigidity to give the accuracy required at the high angular accelerations that occur in the problem. So far as I know, no such flight table has yet been satisfactorily designed, and it therefore appears that one of the most important reasons for using an analogue computer has gone by default.

Mr. W. R. Thomas: Mr. Coales has again raised the question of analogue versus digital for a machine such as Tridac. I still maintain that it is necessary to make it an analogue machine.

This is not solely a computing problem but a physical one. The use of a machine of this type is to enable the designers to understand the performance and limitations of an aircraft or missile and so facilitate the design of a control system that will achieve the results required. It is much easier to understand a problem if we can observe its nature and manner of occurrence on an actual time scale as we can with an analogue machine. In fact we are a fair way towards solving the problem. The programming of a digital computer to solve a 3-dimensional problem of the complexity that Tridac can tackle would be an enormous job, and even now, over six years later, we could not guarantee operation on a 1 : 1 time scale, which is essential in order to include actual components of the system.

The only part of Tridac where the accuracy of a digital machine would be helpful is in the solution of the kinematics, and this is a small part of the whole. A 3-dimensional simulator does not have to be of the size of Tridac; if the assumptions of constant velocity and linear aerodynamics are made, the machine would be one-third of its present size and the power consumption one-fourth. But many problems that the machine is going to investigate could never be tackled.

I consider that the system of Tridac is still sound, but inevitably it takes time to design, build and install a machine of this size, and so the techniques and engineering that have gone into it cannot be up to date. There are better analogue units now available. I do not agree with Mr. Coales's comment that the drift of a d.c. amplifier is still a limit to the accuracy of the analogue computer. Available methods of drift compensation have reduced it to such a level that the main limit is the stability of basic components such as resistors, condensers, etc. Cheaper and more reliable 'choppers' are now available which would make it unnecessary to use the magnetic-modulator technique.

A unit which has been a limitation in analogue computers is the electronic multiplier. The step multiplier, which has been developed to use diodes, offers great hope of providing an accuracy better than 0·1% on a 1 : 1 time scale.

Mr. C. A. A. Wass: The authors state that one of the major factors in the design of Tridac was the desirability of using the real time scale. This facility is valuable but not overwhelming, and I believe that it would have been worth while to forgo this facility if some compensating benefit could have been gained, such as a decrease in size or complexity.

This might lead one to suggest that a digital computer could have been smaller, cheaper and certainly more accurate. Unfortunately, experience shows that the cost of these benefits in terms of reduction of computing speed is surprisingly large. For problems of the kind the authors have suggested, and even some simpler problems, it is found that, with digital computers which have been available in the last few years, the rate of computing is some hundreds of times slower than real time.

Besides excluding any possibility of including real equipment or human operators in the system, this also greatly reduces the rate at which results come out. A series of calculations which might occupy a few hours of real time would engage the slower computers for several weeks.

However, in the future the requirement seems to be, on the one hand, for analogue computers to be somewhat more accurate

and for digital computers to be faster. The accuracy of analogue computers is already limited by the accuracy and stability of resistors, the purity of capacitors, and the stability of supply voltages. It would be unwise to expect any large or sudden improvement in these matters, so that we can only expect gradual improvements in the accuracy of analogue computers as a result of improvement in components, stabilizing of voltages, etc.

With digital computers, however, there are already signs that current computers are considerably faster than the older ones, and there seems to be no fundamental reason why the speed should not go on increasing.

Therefore, in the future the designer will have a more difficult task in deciding between analogue and digital computers in view of the probable large increase in speed of digital computers and the only marginal increase in the accuracy of analogue computers.

Mr. E. G. C. Burt: At the end of Section 3.2, the authors mention that further axis transformation is required to determine the correct control-surface deflections. This resolution has to be done in the missile as well, and it contrasts with the axis resolution as shown in Fig. 4; these kinematic transformations are necessary to account for the fact that the Tridac building is not pursuing a tortuous path through space. The resolution must be done accurately, since any errors present have no counterpart in the real situation, and may lead to serious discrepancies. On the other hand, the control type of transformation is also carried out in the missile, so that errors in the simulator represent exactly corresponding errors in the missile system itself. In fact, Tridac can be used to determine just what tolerances and simplifications are allowable in this area without affecting the overall performance of the system.

A problem, before it is satisfactorily disposed of, goes through a number of stages. By 'problem' I do not mean solving a set of equations, but rather the best way to achieve a given end point, part of the problem being to find out just how to do this. The first stage is the search for possible methods and systems, which is mainly a paper study, using as many approximations and simplifications as are necessary to make the analysis tractable. The next stage is the investigation, probably on a small simulator, of the effects of the neglected terms. This may be followed by an adjustment of the theory in the light of these results. There will often be several such adjustments, as more effects are introduced.

When the small simulator is overloaded, the problem is ready for Tridac; and by this time the staff working on the problem should have a good understanding of it and be able to estimate what kind of additional information is required.

The final output of the machine is not a collection of numbers and graphs. It is an appreciation of the influence of the various factors, with methods of predicting the general system behaviour from the particular cases studied on the simulator. Thus Tridac should be considered not in isolation, but as part of a theoretical-cum-simulator attack on a problem in which the smaller computers—analogue and digital—play their part.

Mr. L. E. McAllister: The primary function of Tridac is to assist aerodynamic design, and the paper has reopened the argument whether such a computer should be of the analogue or digital type. Mr. Coales made a point in favour of analogue devices, when he stated that the operator is given a physical picture of what he is doing.

The organization with which I am associated builds servo systems for the heavier sections of industry, and we have recently constructed an analogue type of servo simulator. This comprises a number of 'building bricks', each of which is a direct analogue of a piece of equipment, and from which a physical representation of the system being studied can be constructed. Behaviour under various conditions, including the effects of changing parameters, is immediately apparent. This direct approach, coupled with the speed with which a problem may be set up and studied, provides the servo-system designers who use the machine with a powerful tool. The programming time and necessary knowledge of techniques would discourage the use of a digital computer for most problems. In this respect, the analogue computer has distinct advantages over the digital computer.

With regard to d.c. amplifier drift, the authors have chosen electro-mechanical correctors for long-term stability. We do not need the accuracy demanded of Tridac, and all our amplifiers are of the normal direct-coupled negative-feedback type. By means of careful circuit design, we have achieved a drift characteristic which approaches the figure quoted in the paper for periods of the order of half an hour, which is long enough for most of our purposes.

[The authors' reply to the above discussion will be found on page 394.]

NORTH-WESTERN CENTRE, AT MANCHESTER, 6TH DECEMBER, 1955

Mr. J. G. Miles: The great complexity of Tridac will naturally give rise to certain queries among engineers who have occasion to use more modest equipments of this type. Several of these queries will be concerned with accuracy considerations. The authors understandably do not commit themselves to a figure for overall accuracy, but it would be interesting to know whether any estimates have been made, and whether they could be checked against practical results in any way. The effect of the stray capacitance of interconnections, etc., appears unlikely to be troublesome in view of the low computing frequencies employed and the fact that many of these interconnections are permanent and can presumably be compensated for. The authors do not mention whether any special measures have been taken to counteract this effect.

In view of the high component accuracies specified, it is somewhat surprising to find saturation effects represented in Fig. 11 merely by limiters, and I wonder whether a closer approximation to the actual saturation curve would not have been justified.

The accuracy of the electronic multiplier unit is given as 0·2%. This figure means little unless related to a frequency range. Is this accuracy obtained when both input quantities are of approximately the same magnitude?

It might be expected that the maintenance of a computer of this complexity would be a major problem, and I would like more information on this point. The authors do not mention the testing of individual electronic computer units, other than checks of power supplies, drift and overloads. How are the input and feedback circuits of these units checked, and is some form of test equation used?

In the actual operation of the computer, I would expect that variable scaling for individual problems would present considerable difficulties in view of the complicated computing circuits involved. A feature of all large-scale special-purpose computers is that the time spent in preparing any particular problem is usually much longer than the time spent in using the computer, and it would be of interest to know the relative number of man-hours spent in programming, setting-up and computing for typical problems, and also the proportional time taken for maintenance.

The authors mention the possible use of high-speed digital

quipment for real-time simulation of this type of problem. It s unfortunate that the part of the computation which appears nost suitable for digital treatment, namely the axis trans-ormations, would presumably still have to be done by analogue neans in view of the real-time requirement, and I wonder whether the effective increase in accuracy would be sufficiently aluable to justify the complex programming involved.

Mr. P. L. Taylor: When such a large sum of money is spent on esearch and development it often happens that interesting new echniques are developed and become generally available as a welcome by-product. Did this happen with Tridac? I would ike to know more about the hydraulic servo-mechanism units.

If asked to tackle the problem again, what changes, if any, would the authors make? For example, the missile is allowed complete rotational freedom, thus adding to the complexity of the co-ordinate transformation system, which accounts for much of the computation required. Has this facility been really necessary in the solution of practical problems? In the aero-dynamic section of the computer, is any distinction made between wind axes and body axes?

Regarded as a whole, the computer is a closed-loop zero-seeking system where the 'zero' is coincidence of target and missile. Thus presumably it will automatically tend to compensate for computational inaccuracy (which might be thought of as computational 'noise'). The earlier in a run the noise occurs, the less it will affect the final result, and so it would seem that computational noise should be important only during the final few seconds of flight. Perhaps the crucial question is whether the authors find that computational noise is less significant than the aerodynamic noise to which an actual missile is subjected.

In Fig. 10, should not a reference second-harmonic voltage be injected into the phase-sensitive detector against which the detector can compare the phase of the amplified modulated drift voltage?

Mr. J. A. Bottomley: Drift-corrected d.c. amplifiers are complex pieces of equipment, giving, with careful design, extremely accurate results. Although such amplifiers are desirable, successful analogue computing can be executed, particularly on smaller machines, using much simpler amplifiers.

In my experience, a straightforward 3-stage d.c. amplifier with a cathode-follower output stage, and an input stage giving some form of compensation for heater drift, suffices in most cases for summation, scale changing, and for the longer-time-constant integrators. Provided that the valve heaters are run continuously (constant-temperature operation), the drift in such an amplifier rarely exceeds 5 mV, referred to the amplifier input, over a period of several hours, and this figure is comparable with that obtained by the use of magnetic modulators. Periodic manual zero-setting is necessary, however, in these amplifiers, whereas with the magnetic-modulator correcting system, a continuous zero-setting is provided. Although manual zero-setting is not practicable in a large machine like Tridac, it does not present much difficulty in smaller machines.

The authors mention the necessity for keeping the grid current in the first valve below 10^{-11} amp. Grid current in an integrator produces an error in the time-constant of integration. A similar error is produced by errors in the measurement of C and R or in the CR product. The accurate measurement of the CR product is not an easy matter, and these constants are, I believe, measured, in the case of Tridac, by a natural-frequency method using two integrators in series, and this in itself is a troublesome task. In my experience, in the majority of cases, the predominent error is in the measurement of C and R and not that arising from grid current, although in our amplifiers the grid current in the first valve is about 10^{-10} amp.

Another point of interest is the hum level in d.c. amplifiers arising from magnetic fields set up by the a.c. supply to the valve heaters. The hum level is often of the order of 5 mV (d.a.p.) in a 1 : 1 amplifier. In a drift-corrected d.c. amplifier, the bandwidth of the a.c. channel is only a few cycles per second. Consequently no reduction of hum is attained by the introduction of drift correction. The hum in the system may completely offset other improvements gained by the use of the drift-correction method. Tridac overcomes such noise by the use of stabilized d.c. heater supplies. Stabilization of 6-volt d.c. supplies is extremely difficult, and such equipment is not always essential, or desirable, where expense is a prime factor.

In my opinion, therefore, for general computing purposes, simpler amplifiers can be used. Larger grid currents and drifts can be tolerated, and periodic manual zero-setting does not present much difficulty. At the same time, the overall accuracy of the computer, although obviously not equal to that of the more complex machine, is sufficient to meet the requirements of most computational problems.

[The authors' reply to the above discussion will be found overleaf.]

NORTHERN IRELAND CENTRE, AT BELFAST, 14TH FEBRUARY, 1956

Mr. R. J. A. Paul: The decision to build such a machine could probably be justified from various considerations, but two main aspects on which I would welcome the authors' comment are the order of accuracy expected and the utilization factor (i.e. ratio of useful computing time to total time involved in setting-up, checking and maintenance). The accuracy cannot obviously be quoted as a definite figure, since it will depend on each particular problem. One application is the simulation of a guided missile attacking an aircraft target, and it would be interesting to know the order of accuracy of 'miss distance' figures. Have the authors any experience on a problem of this magnitude?

It is stated in the paper that, in general, the maximum output error in each computing unit does not exceed 0·1% of the maximum output voltage of 30 volts. However, the maximum error in resistance values is quoted as 0·1%, so that in a resistive feedback amplifier the error could be 0·2% due to this effect alone. With reference to the equation giving the relationship between V_{02} and V_d, if we assume that $Z_3 = Z_2$ and we take the value quoted for β of 1 000, the equation becomes

$$V_{02} = 10^{-3} V_d \left(1 + 2\frac{Z_1}{Z_2}\right)$$

Taking the value of output drift quoted at 1 mV, when the overall gain is unity

$$1 \times 10^{-3} = 10^{-3} V_d \times 3$$

$$V_d = \tfrac{1}{3} \text{ volt}$$

Assuming this value for V_d and considering the case of an integrator of time-constant τ, it may be shown that for a maximum drift error of 0·03 volt, the ratio t/τ must not exceed 45 (where t is the computation time). Thus if $\tau = 1$ sec, the computing time could not exceed 45 sec. Is the computing period of this order?

It is emphasized that this is a drift error of 0·1%, and does not include the errors due to input grid current of the amplifier, finite amplifier bandwidth, finite amplifier gain and absorption change of the capacitor dielectric.

Moreover, an output error of 3 mV is regarded as a fault,

so that presumably three times the above value of error could be expected in the case taken of 45 sec computing time.

The effect of the capacitance loading of the interconnecting cables on the frequency response of the amplifiers and as a possible cause of parasitic oscillations is not mentioned. This effect is surely as serious as that of the cable resistance.

In Fig. 3 the coupling between the main units is fixed and limits, to a large degree, the types of problem which can be solved. Might not a more flexible method of connections increase the scope of the machine?

Mr. C. Snowdon: This question of accuracy really falls into two parts—the first concerns the commissioning tests, and the second concerns the day-to-day accuracy.

In the first instance, it would appear sensible to present to the computer the most complex differential equation which can be solved by conventional methods. A comparison of the two solutions could indicate the accuracy of the computer. Such a sum is quite easy to the computer, whose sole purpose is to solve fantastically complex equations. It would seem, therefore, that the validity of the computer solution is still unproved, and field experience alone could verify the validity of the solution.

On the second point, and assuming that the computer, when first installed, had given the correct answer to any problem presented to it, it may not continue to do so, owing to wear and tear and the limited life of some of the components used within the computer.

The authors indicate that tests on amplifier drift and overloading are sufficient to show that all is well within the machine. I would query this, and suggest that if some of the resistors changed their value the corresponding scalar change would not be noticed and the consequent solution of the problem would be accepted in blind faith. This faith appears to be somewhat unjustified.

The authors give the impression that Tridac has only been used for ground-based missiles attacking targets in 3-dimensional space. Can Tridac cope with the more difficult problem of air-to-air missiles where the attacking and/or attacked aircraft is controlled by a human operator? If a human operator is included, how realistic is the presentation which is shown to him?

Much has been said by other people about the use of alternating current as against direct current, for computational purposes. If the authors had to build a second Tridac would they, in fact, use alternating current, and if so, why?

THE AUTHORS' REPLY TO THE ABOVE DISCUSSIONS

Lt.-Comdr. F. R. J. Spearman, and **Messrs. J. J. Gait, A. V. Hemingway** and **R. W. Hynes** (*in reply*): Mr. Coales has turned our attention to one of the major debating issues in the computing field, namely the relative merits of analogue and digital techniques. The decision taken in 1948 to make Tridac an analogue machine was, in the light of the requirements, the only possible one. Faced with the same problem to-day, our view is that the decision would be the same. The contributions to the discussion made by Messrs. Thomas, Wass and Burt underline the reasons for this view. The two main points in favour of analogue techniques are:

(i) One of the major objectives to be achieved by the use of a machine of this type is an understanding and appreciation of the problems of controlled flight. Analogues to such problems can achieve this end much more effectively than can numerical results.

(ii) The ability of the analogue machine to work in real time and thus absorb parts of real systems into the framework of the simulation is a most valuable property.

The fact that the accuracy of analogue machines is now being limited by the accuracy of standard electronic components has made the users of such machines consider whether digital machines, with their better inherent accuracy, could be applied to their problems. There are two difficulties:

(i) The detailed programming required for a digital machine and its whole mode of operation are such that little physical understanding and appreciation of the problem is obtained from the use of such a machine.

(ii) The present-day digital machines are such that real-time-scale operation, for the Tridac type of problem, cannot be achieved, and even if the necessary very high rate of operation could be achieved, the problems of digital-to-analogue and analogue-to-digital conversion would still have to be solved if parts of real systems were to be incorporated.

Looking well to the future, it may be that the best of both analogue and digital techniques could be merged to give a real-time-scale machine of very high accuracy. This could be achieved if the basic units for addition, integration, etc., could be operated digitally, and thereby accurately, with such units interconnected, not by programming of the type now used for digital machines but by the methods of direct patching as now used on analogue machines. This would preserve the desirable analogue properties but give the accuracy of the digital methods.

Messrs. McAllister and Bottomley both point out that simpler electronic techniques are applicable to simpler and smaller machines. This is clearly true, and much of the complication of the Tridac electronic equipment arises simply from the fact that the number of manual operations, checks, etc., must be kept to a minimum if the machine is to be operable. Also, as the number of computing units is increased, so must the accuracy per unit be increased if the error in the overall simulation is to be kept small.

Messrs. Miles, Paul and Snowdon have raised questions on accuracy, checking and utilization.

The statement that, in general, the maximum output error in each computing unit does not exceed 0·1% of the maximum output voltage of 30 volts, could have been phrased better. It was intended to indicate the general order of accuracy and not to indicate that this precise figure was maintained under all conditions. We agree with Mr. Paul that adverse combinations of resistor tolerances and lengthy integrations with small time-constants can lead to errors greater than 0·1% of full scale. Where necessary, resistors can be selectively matched or trimmed to improve the accuracy, and exceptional care has been taken in the construction and selection of components for the integrators. It is not possible to give a straightforward answer on the question of overall accuracy, since it depends very largely on the nature and size of the problem being solved. For comparison with other analogue machines, an indication of the accuracy of Tridac may, however, be given by quoting the accuracy achieved by the machine in solving a simple closed-loop servo-mechanism problem, such as may be solved by a small analogue machine of about 12 computing amplifiers. The solution consisted of a clamped wave train which the machine reproduced with errors of one part in 700 in frequency and one part in 300 in amplitude. Individual computing units can be thoroughly checked in special test bays. Carefully prepared test problems, with known solutions, can be used for overall checking, not only at the time of commissioning but also on a day-to-day or problem-to-problem basis. When the problem is very large and complex, checking of this type has to be carried out piecemeal. Thus, for example, the aerodynamics response, the kinematics, the control system, etc., are individually checked for correct operation, and the simulation provided by interconnection of these major blocks

f the machine then taken on trust. The final assessment must, f course, be in a comparison of simulator and field results. igures on maintenance times, setting-up times, etc., are not yet vailable, but American experience with large machines of this ype indicates that, of the total machine time, one-third is spent n maintenance, one-third in setting-up and one-third in obtaining olutions.

In answer to Mr. Miles, compensation for voltage drop along ong signal lines is necessary, but no special steps are required o counter inter-wiring capacitance. The frequency characteris- ics of the electronic multipliers are determined simply by the requency characteristics of the feedback amplifiers involved.

In answer to Mr. Taylor, some of the new techniques, developed or Tridac, which have been or could be generally useful are hose associated with drift stabilization, fault-monitoring, hydraulic mechanisms and swash-plate resolution. Once small-angle approximations of the type $\sin\theta = \theta$ are abandoned, and this is essential in Tridac, the only additional complication introduced by considering complete rotational freedom, as compared with limited rotational freedom, is the need for four rather than three synthetic gimbals.

Fig. 10 is correct. The modulator produces a very asymmetric peaky waveform, so that the polarity of the output of the peak detector is dependent on that of the input from the d.c. amplifier.

In answer to Mr. Paul, more flexible methods of interconnecting the main units of Tridac are currently being considered.

In answer to Mr. Snowdon, recent developments in drift-stabilized amplifiers, multipliers, etc., have strengthened the case for having a Tridac type of machine employ direct rather than alternating currents.

DISCUSSION ON

'SOME HALF-TONE CHARGE STORAGE TUBES'*

BEFORE THE RADIO AND TELECOMMUNICATION SECTION, 7TH DECEMBER, 1955

Professor J. D. McGee: With high-capacity pick-up tubes, such as the photo-emissive and photo-conductive types described, why is an electron-multiplier output not used? Might it not give a much longer storage time, since the charges stored on the mosaic are used more effectively?

The authors state that a mica dielectric about 0·000 2–0·000 3 in thick is about the thinnest that can be obtained. Have they considered using evaporated layers of highly insulating materials, such as silicon monoxide?

In the photo-conductive storage tube type VCRX360, is the antimony-trisulphide layer amorphous or crystalline, and why cannot a voltage greater than 10 volts be applied across the layer?

With the same tube, the lag due to the high capacitance of the tube target is relatively short, and the lag that is used in obtaining the duration of a minute or so of the picture is primarily a photo-conductive effect. This would indicate that the lag of the tube might be very dependent on the actual thickness of the photo-conductive layer. Have the authors found any connection between the thickness of the layer and the lag?

This same tube is said to be relatively inefficient for recording and reading flashes of light as applied from a cathode-ray tube displaying radar-type signals. Does this mean that a long-duration flash of low intensity gives a larger response than a short-duration flash of high intensity, the total light flux being the same in both cases? Does the 'reciprocity law', in fact, not hold?

I cannot agree with the statement in Section 4.1 which indicates that the sensitivity of television pick-up tubes is approaching the limit set by the efficiency of presently-available photo-cathode materials. My own calculations indicate that there should be at least two orders of magnitude in sensitivity still to be obtained before we reach the limit set by the shot noise of the primary current from the photo-cathode. I am not taking into account new developments in the efficiency of photo-emissive surfaces, but assume a sensitivity that can readily be achieved at present. If this is taken as 50μA/lumen and it is assumed that the sensitivity of the television camera is limited only by the shot noise inherent in the primary photo-electric emission, for the same optical conditions, the illumination required at the lower limit of practical operation will be about 1% of that required by the most sensitive camera now available.

I should like to know the exact function of electrode 9 in Fig. 9, when the reading gun is being used as the signal electrode. This electrode is stated to reduce the spurious signal, but the explanation seems neither very adequate nor clear.

Mr. L. S. Allard: Undoubtedly the storage surface is the heart of the storage tube, and it would therefore be interesting to have more details of the preparation of these surfaces. Limitations of these surfaces have indeed been mentioned, and, in particular, the problem of spots has been quoted. I infer from the paper that the causes of these spots are dust and dirt particles. In the organization with which I am associated we have been working on other types of storage tube, and have been troubled with similar defects, but we are not convinced that they are all due to dust particles. Various types of storage surface have been investigated, including magnesium fluoride and anodized aluminium, and on a number of these tubes, spots have been noted. Some of these have been connected with surface impurities other than dust or dirt, while others have been connected with impurities located in the base-plate of the target.

Was the choice of calcium fluoride as the storage medium for

* Webley, R. S., Lubszynski, H. G., and Lodge, J. A.: Paper No. 1883 R, July, 1955 (see 102 B, p. 401).

the VCRX350 governed by its better performance as regards spots and shading signals?

Mention is made of the long-term storage effects on the VCRX350, and this prompts the question whether the operation of these tubes is dependent on a simple charging and discharging of condensers. Our own experience with anodized aluminium targets has indicated the retention of charges for periods much longer than that given by calculation, and I wonder whether some volume-storage effect is taking place.

With regard to the mesh used in the VCRX350, it is stated that a better performance is obtained when the storage surface is on the side containing the vertices of the triangles. Is there any fundamental reason for this, and is this observed phenomenon in agreement with simple condenser charging theory?

Has the VCRX326, in a modified form, been used as a direct-viewing storage tube? By removing electrode 9 in Fig. 8, and introducing an accelerating field between mesh 7 and a fluorescent screen on the end of the tube, it would appear that a direct-viewing storage tube might be possible.

Finally, have any of these storage tubes been used in digital computers? In the Manchester University computer, the storage-tube type VCRX266 is used with about 1 000 digits per tube. The limitation on storage capacity is due not to spot-size considerations, but to the redistribution of the secondary electrons when the charge is deposited on the screen. From the figures given in the paper, with a resolution of 700 spots per picture diameter, I imagine that there could be a considerable gain in the capacity for storing digits, or do these tubes also suffer from adverse secondary-electron redistribution effects?

Mr. L. J. Braybrook: Requirements exist for bright radar displays in aerodrome control towers, where it is desirable to use radar equipment and yet see part of the airfield, and in air-traffic control centres, where radar equipment is used in conjunction with procedural control which involves reading and writing in artificial light.

The operational advantages of bright displays using television technique are obtained at the cost of additional equipment and maintenance, which can be considerable when off-centre displays are required for a number of sectors under surveillance from a single radar equipment. The storage-tube characteristics need to be matched to the radar information, and this is especially the case when the radar is susceptible to precipitation clutter; a controllable storage time is then essential. The VCRX350 is superior to the VCRX360 for this purpose, but I should like to know whether it is possible to select the latter tubes from a batch so that the delay time suits particular radar conditions.

An experimental system similar to that described in Section 3.2.10 has been used for radar p.p.i. display; the lens was of 2 in focal length and was operated at an aperture of f2.9; the signals from the VCRX360 were amplified by the video-frequency section of a television console and displayed on a 15 in aluminized cathode-ray tube. This was not an ideal system, but Fig. A indicates the performance with a definition of 625 lines and 50 frames/sec.

The brightness ratio between a target and its background was measured using a contrast photometer both for a normal p.p.i. tube with a magnesium-fluoride phosphor and the VCRX360 display. Measurements were made with and without the appropriate optical filters over the faces of the display tubes. The fluoride tube was superior in semi-darkness, but when the ambient light was sufficient for reading and writing, the television display had the advantage.

Mr. P. L. Waters: With regard to the control erasure time on the VCRX350, with our apparatus at the Radar Research Establishment we find that the actual output signal level decreases when the erasure time decreases. From the paper, it appears

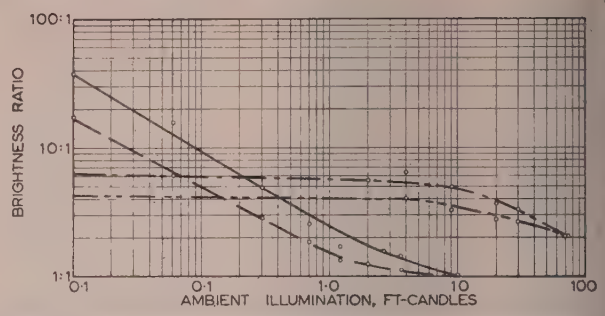


Fig. A.—Brightness ratio of echo to background at various ambient light levels.

Echo: Edge of permanent clutter from North Downs.
Beam width: 1 deg.
Pulse duration: 1 microsec.
Rate of turning: 4 r.p.m.

———— Fluoride display, CV429, with orange red filter.
– – – – Fluoride display, CV429, without filter.
— – — High-intensity display, VCRX360, with neutral filter.
— – – — High-intensity display, VCRX360, without filter.

that, once the electrons have passed through the storage mesh, the erasure mechanism should merely send them back to the insulating material and should not affect the signal strength.

Fig. 12 shows the output from a number of these tubes when set up for a given erasure rate or storage time. Although it has a lag curve superior to that of the VCRX360, I notice that the VCRX350 gives only approximately half the output signal. When first associated with this development of tubes with a storage mesh, I thought that we ought to be able to obtain a larger output signal with the VCRX350 than with the VCRX360 since we expected to be able to adjust the decay rate independently of the output signal current.

We have found that the background uniformity is a more serious trouble in performance than the actual amplifier noise. Some figures for signal/noise ratio are quoted, and I assume that they refer to (peak signal)/(r.m.s. amplifier noise). I wonder whether any information is available on the possibility of an improvement, particularly with the VCRX350. We have found this rather inferior to the VCRX360 on background uniformity although, as has already been pointed out, it has a number of advantages as a tube—notably the control of the lag. Is there some fundamental problem in this mesh-grid type of tube which limits the performance as regards background uniformity or is it a question of controlling the manufacturing processes more carefully?

Mr. P. A. Newson: The chief reference during the discussion has been to the use of the tubes for radar displays, but I would not have thought a half-tone tube particularly valuable for that purpose. I would have thought that one either had the echo or not, and therefore the actual contrast was unimportant, provided that it was adequate. On the other hand, the title of the paper indicates that a projected major use was for television purposes.

The picture shown exhibited very pronounced charge-storage properties, but it was very deficient in contrast. Therefore is the contrast law of the pick-up tube similar to that of a normal camera tube, which is fairly conveniently related to the contrast law of the reproducer, or is it inherently different? Were we merely suffering from the effects of an experimental tube, and perhaps an imperfectly aligned recorder-to-producer chain? Possibly the use of commercial television receivers for these tubes does not improve matters, and transportable apparatus is always liable to irregularities.

Has the possible use of these tubes for television standards onversion been investigated? I believe that this was first done y merely writing on a normal cathode-ray tube with or without ot-wobble or smoothing device, and then re-reading it with a ormal television camera, working on the revised standards. his is a possible way of carrying out the operation, but I do ot know whether the inherent resolution of the mesh would e suitable for such an application.

Mr. A. V. Lord: The authors mention erasure times of the rder of a few seconds. In television there are several important pplications for a storage tube capable of the efficient storage f one television field for the duration of one field and having he important property that effectively complete erasure is btained either during the reading field or during the suppression nterval following reading. I would appreciate the authors' omments as to the feasibility of such a tube.

One of the main problems in television standards conversion s caused by the possible difference in field (and picture) fre- uencies between the input and output standards.* At present, difficulties due to small differences of frequency may be overcome y means of a suitable phosphor persistence in the writing athode-ray tube.† A more satisfactory solution could be btained by means of a writing tube such as that proposed by chroeter and described by Knoll and Kazan.‡ Could the uthors discuss the possibilities of developing from the VCRX326 a tube having similar properties?

Mr. R. S. Webley, Dr. H. G. Lubszynski, and **Mr. J. A. Lodge** *(in reply)*: The use of an electron multiplier, as mentioned by Prof. McGee, would doubtless increase maximum storage time n both the VCRX343 and VCRX360. For the effect in the ormer, we would refer Prof. McGee to Forgue's paper.§ The ffect in the latter tube would be to reduce the noise by a factor f about 3 in the system mentioned, thus extending the reading ime, which, however, was already adequate for the purpose in nind. The additional complication is not considered worth while. Several attempts were made to produce thin quartz argets. In only one instance was a target of sufficient insulation btained. With this the lag was about 1 min, but the back- ground was extremely poor.

We believe the layer in the VCRX360 to be amorphous. The maximum gradient across the layer is limited by excessive dark current. For maximum sensitivity, the light must penetrate right through the layer. This, in turn, limits the maximum thickness of the layer and thus the total voltage. Thicker layers generally show a shorter lag. A long-duration low-intensity flash can give a larger response than a high-intensity one if the duration of the latter is considerably less than the build-up lag time.

Calculations show that, on the basis of the signal/noise ratio generally accepted as giving a high-quality picture, an improvement of 10 : 1 is possible beyond that now available. If one is prepared to degrade the picture appreciably, the possible improvement rapidly approaches the figure quoted by Prof. McGee.

In reply to Mr. Allard, calcium fluoride was found to be the most satisfactory of a large number of layers tried. We have observed no effects which can be ascribed definitely to volume storage, although one or two occasional phenomena could be explained on this basis. We have no explanation for the effect of mesh orientation. In the past, we have considered ideas for directly-viewed tubes along the lines suggested. However, no actual development work has been done. The use of the VCRX326 as a digital store is possible, but has not been tried. A modified VCRX343 has been used with some success. Neither of the other tubes seems suitable.

We are grateful to Mr. Braybrook for his contribution to the discussion. Under suitable manufacturing conditions, selection of VCRX360's to meet definite storage requirements is possible.

The deficit of electrons in the returning beam constitutes the output signal. In the erasure condition some beam is reflected back through the mesh, thus reducing the deficit and hence the signal strength, as has been observed by Mr. Waters. Background uniformity of the VCRX350 has been improved in recent months, so that shading is now unnoticeable in normal operation.

In reply to Messrs. Newson and Lord, a half-tone display is advantageous for radar in detecting weak echoes in the presence of noise, especially when integration can be used to enhance signal/noise ratio. The low contrast in the half-tone picture shown was due to limitations in the portable channel used in the demonstration.

None of the links described is suitable for television standards conversion because of relatively long erasure times. For this purpose, the standard C.P.S. Emitron would be close to the ideal.

Several tubes have been proposed which could, in principle, satisfy Mr. Lord's requirements, among them being the Metrechon. None has, at present, the required erasure speed.

* Zworykin, V. K., and Ramberg, E. G.: 'Standards Conversion of Television ignals', *Electronics*, January, 1952, p. 86.
† Lord, A. V.: 'Conversion of Television Standards', *B.B.C. Quarterly*, 1953, 8, . 108.
‡ Knoll, M., and Kazan, B.: 'Storage Tubes' (John Wiley and Sons, 1952), p. 81.
§ Forgue, S. V.: *RCA Review*, December, 1947, p. 633.

621.316.86 : 621.3.011.21 : 621.3.012.8

Paper No. 1751 M
Dec. 195

THE SPECIFICATION OF THE PROPERTIES OF THE THERMISTOR AS A CIRCUIT ELEMENT IN VERY-LOW-FREQUENCY SYSTEMS

By C. J. N. CANDY, Ph.D., Graduate.

(*The paper was first received* 24*th May, and in revised form* 16*th August*, 1954. *It was published in December*, 1954, *and was read before a Joint Meeting of the* MEASUREMENT AND CONTROL SECTION *and the* RADIO AND TELECOMMUNICATION SECTION 29*th November*, 1955.)

SUMMARY

An analysis based on sinusoidal applied voltages shows that a conductor whose resistance is a function of temperature may be represented by an equivalent circuit having a semicircular impedance locus. An expression for the distortion of the waveform is also obtained, and this is found to be small provided that the alternating current is less than a quarter of the steady polarizing current which flows in the conductor. The impedance loci of a bead-type thermistor are plotted by means of a null technique. A typical impedance varied from a negative resistance of 3 000 ohms at very low frequencies to a pure inductance of 2 000 H at 0·3 c/s and then to a positive resistance of 5 000 ohms at high frequencies. The use of the equivalent circuit is illustrated by designing phase-shift networks suitable for use in the stabilizing of very-low-frequency control systems. These circuits may be used in systems where either a.c. or d.c. data transmission is employed.

LIST OF PRINCIPAL SYMBOLS

p = Instantaneous power.
$\bar{\theta}$ = Mean temperature.
θ_0 = Ambient temperature.
θ = Instantaneous temperature.
K = Dissipation constant.
S = Thermal capacity.
α = Temperature coefficient of resistance.
a, b = Thermistor parameters related by the expression $R = a\varepsilon^{b/\theta}$.
ΔR = Incremental resistance.
$\bar{R}$ = Mean resistance.
R = Instantaneous resistance.
Z_I = Thermistor incremental impedance.
τ = Time-constant.
R_0 = Low-frequency resistance of the thermistor.
R_∞ = High-frequency resistance of the thermistor.
$\mathrm{D} = d/dt$.

(1) INTRODUCTION

The paper is concerned with the investigation and development of circuits employing the thermal time-constants of temperature-sensitive resistors to obtain phase shifts of electrical signals. Large time-constants are easily realizable; with good thermal insulation it is not difficult to obtain a time-constant of several minutes in a physical space of a few cubic centimetres.

The fact that a phase difference exists between the voltage and the current in a temperature-dependent resistor has been demonstrated by a number of writers.[1–10] It is also known that in addition to this phase shift there is considerable harmonic distortion of the waveform. The work described in the paper shows that if the thermally-sensitive resistor is polarized from a steady source of current which is at all times greater than the signal current, the harmonic distortion is negligible and the phase shift is greatly increased. Under these conditions the voltage across the resistor is made up of three major components. These are:

(*a*) The product of the signal current and the mean resistance in phase with the current.

(*b*) The product of the polarizing current and the fundamental component of the incremental resistance, lagging the current by an angle dependent on the thermal time-constant of the resistor.

(*c*) A steady voltage due to the steady polarizing current flowing in the mean resistance.

Voltages (*a*) and (*b*) combine to give greater shift than is possible with an unpolarized resistor.

In order that small changes of signal current in the temperature-sensitive resistor shall cause appreciable changes of resistance it is necessary to use material which has a large temperature coefficient. For this reason the thermistor is a very suitable circuit element, especially since it is available commercially in a variety of forms. It has been convenient to limit the present investigation to the study of thermistors and thermistor circuits although many of the results are applicable to any thermally sensitive resistor such as the filament lamp. The inherent difficulty due to the variation of the impedance with the ambient temperature is discussed in an appendix.

No really satisfactory method for specifying the properties of circuits containing thermally sensitive resistors has been presented in the published literature. An equivalent impedance for such a resistor is derived in the paper, and all the parameters of this impedance are shown to be easily determined by simple tests. The vector locus of this impedance is a semicircle—a very convenient basis for the design of circuits employing thermal time-constants.

The possibility of employing the thermal time-constant of temperature-sensitive resistors in low-frequency circuits has been proposed by a number of writers. In 1943 Tustin[8] suggested the use of thermal time-constants to obtain a phase advance in a signal, in both a.c. and d.c. systems. The need for a steady polarizing current in the resistor was pointed out, and a simple analysis was used to obtain results which are similar to those described in the paper. In 1946 Becker, Green and Pearson published a comprehensive paper on the properties and uses of thermistors. They suggested the possible use of thermistors as circuit elements in very-low-frequency circuits but made no attempt to develop the idea. Cooper and Seymour[6] have made a detailed analysis of the dynamic characteristics of temperature-dependent resistors. They applied their results to the use of a filament lamp in the amplitude control of a 31-c/s oscillator.

The particular case of the thermistor is discussed in detail by Bollman and Kreer.[4] Differential equations for the dynamic properties of thermistors are derived, but these are not in a form that can be readily used as a basis for the design of thermistor circuits. Alternative equations have been derived by Smith,[5] but these are no simpler than those given by Bollman and Kreer. Smith discussed the possibility of using thermistors in low-frequency filters and in modulators, but made

Dr. Candy is now with Messrs. S. Smith and Sons, Ltd., and was formerly at the University College of North Wales, Bangor.

little headway in the direction of a simple design method. Most of the discussion was based on an equivalent circuit consisting of a series *LR* circuit; it was shown that both the inductance and the resistance are functions of frequency, as would be expected, since the complete circuit must include a resistor shunting the inductance. An interesting ultra-low-frequency oscillator was described, involving a circuit due to Stone and described fully in Reference 3. The frequency of oscillation is fixed by a resonant circuit consisting of a parallel combination of a thermistor and a very large capacitor.

It should be added that a number of writers[2] have calculated the phase angle between the voltage and current in a metal filament lamp, and also the extent of the harmonic distortion of the waveform due to the variation in the resistance over each cycle. These results are used to determine the unwanted voltage in the indicator system of a voltage stabilizer, which consists of a bridge circuit composed of two resistors and two lamps.

(2) THE PHYSICAL PROPERTIES OF THERMISTORS

A thermistor[1–5] is a circuit element whose electrical resistance varies over a very wide range with changes in temperature. In contrast to metals, which have small positive temperature coefficients of resistance, the semi-conductors from which thermistors are made have relatively large negative temperature coefficients.

Thermistors are made in a variety of forms; the resistance material may be formed into small bead-type elements, pressed into discs or extruded into rods. A typical thermistor (type A1522/100*) consists of a small bead fused to two very fine terminal wires and mounted in an evacuated glass capsule. Heat losses are thus reduced to a minimum, and the change in temperature resulting from a given change of current is high.

(2.1) The Static Characteristic

One of the most interesting and useful properties of a thermistor is the way in which the voltage across it changes as the current

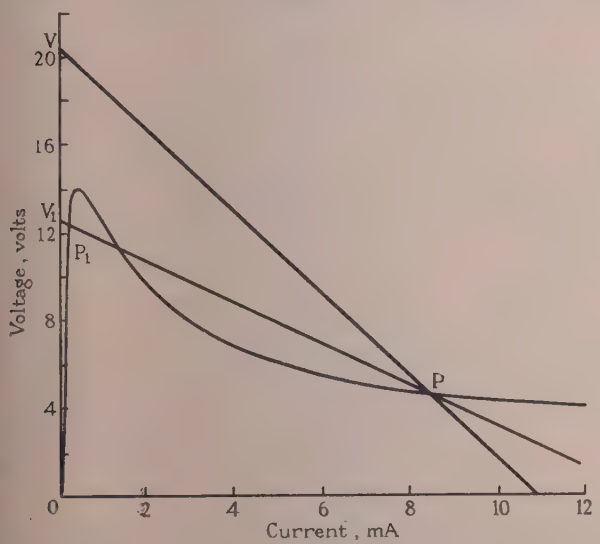


Fig. 1.—Load lines on thermistor characteristics.

increases. The curve in Fig. 1 shows this relation for a bead-type A1522/100 thermistor. In plotting the curve, sufficient time is allowed after each change of current for the voltage to attain a new steady value; hence the curve is called the static characteristic. For sufficiently small currents, the power dissipated is too small to heat the thermistor appreciably, and the voltage is proportional to the current. However, as the current reaches larger values the power increases, the temperature rises above the ambient temperature, the resistance decreases and therefore the voltage is less than it would have been had the resistance remained constant. At some particular current value the voltage attains a maximum or peak value. Beyond this point an increase in current causes a decrease in voltage, and the thermistor has a negative resistance to incremental currents. For this reason a thermistor should never be connected directly to a source of voltage with a low internal impedance, since this would result in excessive current and damage to the thermistor.

* Standard Telephones and Cables, Ltd.

(2.2) Relation between Resistance and Temperature

The resistance of a thermistor, R, decreases with increasing temperature according to the approximate expression

$$R = a\varepsilon^{b/\theta} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where R is the resistance in ohms at a temperature of θ degrees K. The quantities a and b are fundamental parameters of the thermistor.

If a thermistor has a resistance $\bar{R}$ at a temperature $\bar{\theta}$ then

$$R = \bar{R}\varepsilon^{b(1/\theta - 1/\bar{\theta})} \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

For very small changes of temperature, such that $\bar{\theta} \simeq \theta$ and $b(1/\theta - 1/\bar{\theta}) \ll 1$,

$$R \simeq \bar{R}\left[1 - \frac{b}{\bar{\theta}^2}(\theta - \bar{\theta})\right] \quad . \quad . \quad . \quad . \quad (3)$$

Therefore the temperature coefficient of resistance of a thermistor for small variations about a mean temperature $\bar{\theta}$ is

$$\alpha = -\frac{b}{\bar{\theta}^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

(3) THE STEADY-STATE PROPERTIES OF THERMISTOR CIRCUITS

It is stated in Section 2.1 that a thermistor should never be connected directly to a voltage source of low internal impedance; a resistor should be placed in series with the thermistor. The value of this resistor, and the voltage, must be so chosen that the current in the circuit has a certain convenient value.

In Fig. 1, let the point P denote the required quiescent conditions, and the point V the applied voltage. Then the series resistance must be made equal to the slope of VP. If, however, the voltage were lower, as denoted by V_1, and the resistance were made equal to the slope of V_1P, the steady-state conditions arrived at after switching on would be those indicated by P_1 and not by P. That is, the thermistor would not be biased over the hump, unless some external help such as a pulse of voltage were given or some other means of warming the thermistor were used.

These artificial methods are not recommended, because there is a danger of the circuit conditions returning to P_1 when a disturbance is superimposed on the steady current. For example, assume the current in the thermistor circuit shown in Fig. 2 is

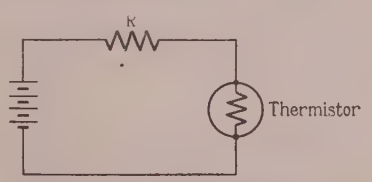


Fig. 2.—Basic thermistor circuit.

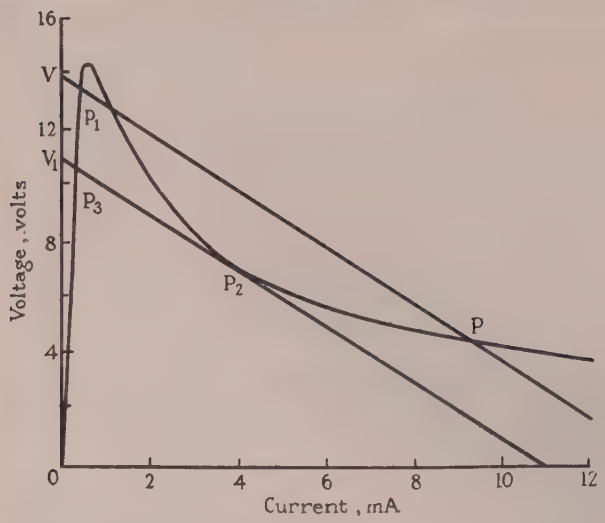


Fig. 3.—Instability of the operating point.

given by P in Fig. 3. If the voltage is reduced the current in the circuit will decrease, and the operating point P will move to the left. Any attempt to reduce the voltage below the point V_1 will result in the operating point "triggering" from P_2 to P_3, and when the voltage is again increased to its original value, the steady-state conditions will be defined by P_1, not P.

(4) IMPEDANCE OF A CONDUCTOR WHOSE RESISTANCE IS A FUNCTION OF TEMPERATURE

The static characteristics described in Section 2.1 are obtained when the thermistor is in thermal equilibrium with its surroundings. If the thermistor is not allowed to settle to thermal equilibrium as each point is plotted, but has its current continuously varied, the voltage across it will depend upon whether the current is being increased or decreased. A continuously increasing current would produce voltages above those of the static characteristic, while a decreasing current would have the opposite effect. Thus the voltage contains a component which depends on the rate of change of current. This component appears because the temperature of the thermistor and the voltage across it lag behind the changes in the power, and it is only apparent when the currents are large enough to heat the thermistor appreciably.

It has been shown that the passage of a very small current has little effect on the thermistor resistance; Ohm's law is obeyed near the origin of the static characteristic. If the thermistor is polarized with a steady current, however, and the small variable current superimposed, appreciable change of resistance can occur, and the passage of the polarizing current through this variable resistance produces a variable voltage drop across the thermistor in addition to that due to the variable current.

(4.1) Impedance of a Polarized Conductor

The impedance of a conductor may be determined by assuming that a given current flows in the conductor, and calculating the voltage across it due to this current.

Consider a conductor carrying a direct polarizing current, i, and let there be superimposed on this an alternating current

$$i_1 = \hat{I}\cos\omega t \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

In order to calculate the voltage, we have the incremental resistance

$$\Delta R = \alpha\bar{R}(\theta - \bar{\theta}) \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Assume that the peak value of the alternating current, $\hat{I}$, is small, so that the variation of temperature $(\theta - \bar{\theta})$ is also small and $\Delta R \ll \bar{R}$. For small changes, the resistance is given by

$$R = \bar{R}[1 + \alpha(\theta - \bar{\theta})] = \bar{R} + \Delta R \quad . \quad . \quad . \quad (7)$$

and the instantaneous power dissipated in the conductor is

$$p = R(i + \hat{I}\cos\omega t)^2 \simeq \bar{R}(i^2 + 2i\hat{I}\cos\omega t + \tfrac{1}{2}\hat{I}^2 + \tfrac{1}{2}\hat{I}^2\cos 2\omega t) \quad . \quad . \quad . \quad . \quad (8)$$

The power supplied must equal the rate of increase in the internal energy plus the rate of loss of heat, so that

$$p = S\frac{d\theta}{dt} + K(\theta - \theta_0) \quad . \quad . \quad . \quad . \quad . \quad (9)$$

Therefore
$$p - K(\bar{\theta} - \theta_0) = S\frac{d\theta}{dt} + K(\theta - \bar{\theta})$$
$$= \frac{S}{\alpha\bar{R}}\frac{d(\Delta R)}{dt} + \frac{K(\Delta R)}{\alpha\bar{R}} \quad . \quad . \quad (10)$$

In the steady state the mean power supplied must equal the mean power dissipated, so that

$$\bar{R}\left(i^2 + \frac{\hat{I}^2}{2}\right) = K(\bar{\theta} - \theta_0) \quad . \quad . \quad . \quad . \quad (11)$$

Eliminating p and θ from eqns. (8), (10) and (11), and writing $D = d/dt$, we have

$$(K + SD)\Delta R = 2i\hat{I}\bar{R}^2\alpha\cos\omega t + \tfrac{1}{2}\hat{I}^2\bar{R}^2\alpha\cos 2\omega t$$

The steady-state solution (particular integral) of this differential equation is

$$\Delta R = 2i\hat{I}\bar{R}^2\frac{\alpha}{K}\cos\phi_1\cos(\omega t - \phi_1)$$
$$+ \hat{I}^2\bar{R}^2\frac{\alpha}{2K}\cos\phi_2\cos(2\omega t - \phi_2) \quad . \quad (12)$$

where $\phi_1 = \arctan\dfrac{\omega S}{K}$, and $\phi_2 = \arctan\dfrac{2\omega S}{K}$

The voltage across the conductor is then given by

$$v = (i + \hat{I}\cos\omega t)(\bar{R} + \Delta R)$$

from which

$$v = \hat{I}\bar{R}\cos\omega t + 2\hat{I}(\bar{R}i)^2\frac{\alpha}{K}\cos\phi_1\cos(\omega t - \phi_1)$$
$$+ \hat{I}(\bar{R}\hat{I})^2\frac{\alpha}{4K}\cos\phi_2\cos(\omega t - \phi_2)$$
$$+ \frac{i(\bar{R}\hat{I})^2\alpha}{2K}[2\cos\phi_1\cos(2\omega t - \phi_1) + \cos\phi_2\cos(2\omega t - \phi_2)]$$
$$+ \frac{\hat{I}(\bar{R}\hat{I})^2\alpha}{4K}\cos\phi_2\cos(3\omega t - \phi_3)$$
$$+ \bar{R}i + \frac{i(\bar{R}\hat{I})^2K\alpha}{K^2 + \omega^2S^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

Thus the voltage consists of a component of the same frequenc as the current but out of phase with it, and components of twic and three times this frequency, together with a direct voltage.

When the peak alternating current $\hat{I}$ is very small compared with the polarizing current i

$$v \simeq \hat{I}\bar{R}\cos\omega t + \frac{2\hat{I}(\bar{R}i)^2\alpha}{\sqrt{(K^2+\omega^2S^2)}}\cos(\omega t - \phi_1) + \bar{R}i \quad . \quad (14)$$

The harmonic distortion of the voltage waveform will then be negligible. The incremental voltage across the conductor may be expressed in vectorial form thus

$$V = \left[\bar{R} + \frac{2(\bar{R}i)^2\alpha}{K+SD}\right]I \quad . \quad . \quad . \quad . \quad (15)$$

where V and I are the voltage and current vectors respectively. The term $(\bar{R}i)$ represents the steady voltage due to the polarizing current, and as such is not included in the incremental voltage.

Eqn. (15) indicates that the polarized conductor may be regarded as having an equivalent impedance to alternating currents, given by

$$Z = \left[\bar{R} + \frac{2(\bar{R}i)^2\alpha}{K+SD}\right] \quad . \quad . \quad . \quad . \quad (16)$$

It should be noted that Z is a function of the polarizing current, because the expression includes i, and also because $\bar{R}$ and α will themselves be functions of i.

(4.2) The Unpolarized Conductor

An expression for the voltage across an unpolarized conductor may be obtained by putting i equal to zero in eqn. (13). Then

$$v = \hat{I}\left\{\bar{R}\cos\omega t + \frac{(\bar{R}\hat{I})^2\alpha}{4\sqrt{(K^2+4\omega^2S^2)}}\right.$$
$$\left.[\cos(\omega t - \phi_2) + \cos(3\omega t - \phi_2)]\right\} \quad . \quad (17)$$

The analysis assumes that the incremental resistance is small compared with $\bar{R}$, so that

$$\frac{(\bar{R}\hat{I})^2\alpha}{2\sqrt{(K^2+4\omega^2S^2)}} \ll \bar{R}$$

This will hold for small currents, under which conditions the component of voltage that is frequency-dependent will be small. Furthermore, the voltage is not directly proportional to the current; therefore the unpolarized conductor may not be represented by a linear impedance.

The following Sections will deal only with polarized conductors having characteristics approximately linear for small currents, and capable of introducing large phase shifts between the voltage and the current.

(5) EQUIVALENT IMPEDANCE OF A POLARIZED THERMISTOR

The temperature coefficient of a thermistor for small changes is given in eqn. (4). This may be substituted in the expression for the impedance of a general conductor in eqn. (16) to give the impedance Z_T of a polarized thermistor:

$$Z_T = \bar{R} - \frac{2b\bar{R}^2i^2}{\theta^2(K+DS)} \quad . \quad . \quad . \quad . \quad (18)$$

In order that this expression may be simplified, consider the thermistor to be polarized so that the operating point is defined by P in Fig. 4. If an alternating current, of small amplitude and very low frequency, is superimposed on this polarizing current, the voltage for any current will be given by the corresponding ordinate of the static characteristic. When the frequency of the impressed current is high, however, the resistance will not change with individual cyclic changes of current; it will assume an

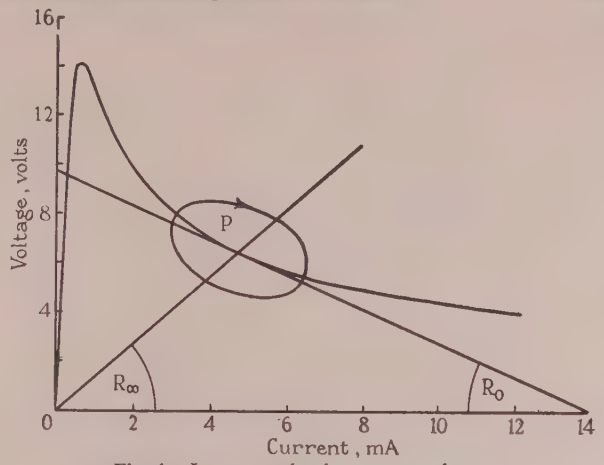


Fig. 4.—Incremental voltage-current locus.

average value equal to the slope of OP, provided that the amplitude of the alternating current is small compared with the polarizing current.

It may be concluded that the incremental low-frequency impedance of the thermistor is a resistance equal to the slope of the thermistor static characteristic at the operating point P. This resistance is negative, and will be denoted by $-R_0$. The high-frequency impedance is a positive resistance, R_∞, equal to the slope of OP. It should be noted that with this definition both R_0 and R_∞ are positive over the greater part of the static characteristic.

The expression for the thermistor impedance may be written in terms of R_0 and R_∞. A convenient form is

$$Z_T = R_\infty - \frac{R_\infty + R_0}{1 + D\tau} \quad . \quad . \quad . \quad . \quad (19)$$

where τ $(= S/K)$ is the time-constant of the thermistor.

The fact that the reactive component of the thermistor impedance is positive may be demonstrated by considering the voltage across the thermistor when the frequency of the impressed current has a value intermediate between the two extreme values considered above. Under this condition the voltage will not be given by the static characteristic, because the change in the resistance will lag behind the changes in the current, i.e. for a given increasing current the resistance, and therefore the voltage, will be greater than the steady value, and for decreasing currents it will be less. The operating curve will therefore have an oval shape as shown in Fig. 4, and the direction of rotation will be clockwise, showing that the voltage leads the current.

The numerical value of the thermistor impedance for any given polarizing current may be obtained by measuring R_0 and R_∞ on the static characteristic, and determining the time-constant by a simple experiment in which the transient response of a thermistor to a sudden increase in the polarizing current is studied. In Fig. 5A the resistance R is very much greater than

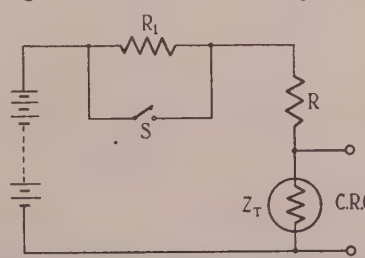


Fig. 5A.—Circuit for obtaining transient response.

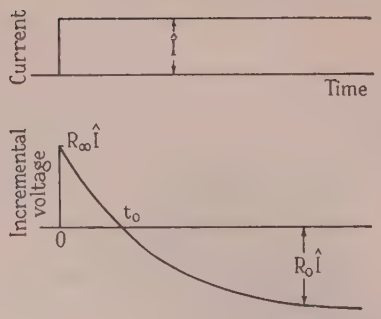


Fig. 5B.—Transient response.

R_0 and R_∞, so that the thermistor may be assumed to be fed from a constant-current source.

The voltage appearing across the thermistor, v, due to a step in the current, is determined by means of Laplace transforms:

$$v = \Gamma^{-1}\left(R_\infty - \frac{R_\infty + R_0}{1 + s\tau}\right)\frac{1}{s}\hat{I} = [R_\infty - (R_\infty + R_0)(1 - \varepsilon^{-t/\tau})]\hat{I} \quad . \quad . \quad . \quad . \quad (20)$$

This is of the form shown in Fig. 5B. The incremental voltage first rises and then falls to a steady negative value. The incremental voltage is zero at time t_0 such that

$$R_0 = (R_\infty + R_0)\,\varepsilon^{-t_0/\tau}$$

Then

$$\tau = \frac{t_0}{\log_\varepsilon\left(1 + \frac{R_\infty}{R_0}\right)} \quad . \quad . \quad . \quad . \quad (21)$$

The time t_0 may be measured either by calibrating the time-base of the oscillograph with a 50-c/s ripple, or directly with a stop-watch for thermistors with large time-constants. Some typical values are as follows:

1·4sec for thermistor type A1522/100.
4·4sec for thermistor type B5412/60.
32sec for thermistor type K1331/120.

In all cases the time-constants were found to be approximately independent of the value of the polarizing current.

(5.1) The Equivalent Circuit and Locus of the Thermistor Impedance

The expression for the thermistor impedance given in eqn. (18) suggests two possible equivalent circuits for the thermistor. These are shown in Fig. 6, and some typical values of the circuit

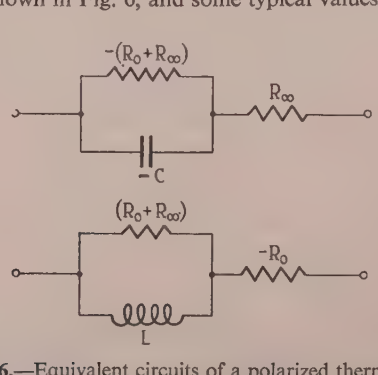


Fig. 6.—Equivalent circuits of a polarized thermistor.

Table 1

IMPEDANCE OF THERMISTOR A1522/100 WITH VARIOUS POLARIZING CURRENTS

i	R_∞	R_0	L	C
mA	ohms	ohms	H	μF
13	300	130	600	3 250
8	650	350	1 400	1 400
5	1 300	625	2 700	720
1	12 200	4 000	22 500	86

elements for the bead-type thermistor A1522/100 are given in Table 1 for various polarizing currents.

The locus of the impedance as the frequency varies will be a semicircle of the form shown in Fig. 7, where OQ represents the

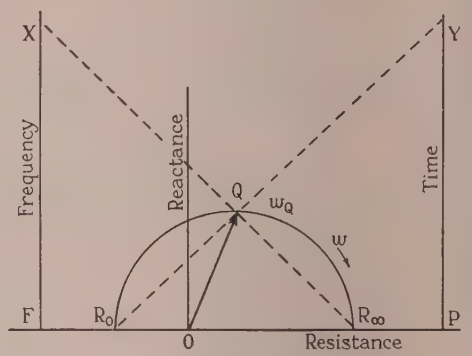


Fig. 7.—Locus of thermistor impedance.

value of the impedance at any angular frequency ω_Q. Ther XF will be proportional to ω_Q, and PY will be proportional tc $1/\omega_Q$. The impedance varies from a negative resistance $-R_i$ at very low frequencies to a pure reactance $\sqrt{(R_0 R_\infty)}$ wher $\omega = (1/\tau)\sqrt{(R_0/R_\infty)}$, and finally to a positive resistance R_∞ a high frequencies.

(6) A METHOD OF OBTAINING THE THERMISTOR-IMPEDANCE LOCUS EXPERIMENTALLY

The impedance locus may be plotted by means of the circuit[1] shown in Fig. 8. The signal current is obtained from a trans mitter magslip driven at constant speed by a Velodyne system the modulated voltage obtained from the magslip stator i rectified and applied to the thermistor network. Two thermistor are used in a balanced circuit, in order to eliminate the stead direct voltage due to the polarizing current, which is supplied b the battery and the resistors R′.

A monitoring voltage is taken from the magslip stator throug a calibrated phase shifter. This voltage is rectified and the applied to the X-plates of the oscillograph.

At each test frequency the potentiometer, n, is adjusted unti the vertical deflection of the oscillograph trace is independent o the position of the switch S. With the switch in position 1, th phase shifter in the monitor circuit is adjusted until the elliptica Lissajous figure degenerates into a straight line, and a readin of the phase-shifter position is taken. The switch is then place in position 2, and the phase shifter readjusted and read; th difference between the readings, ϕ, gives the phase angle betwee the voltages v_T and v_n.

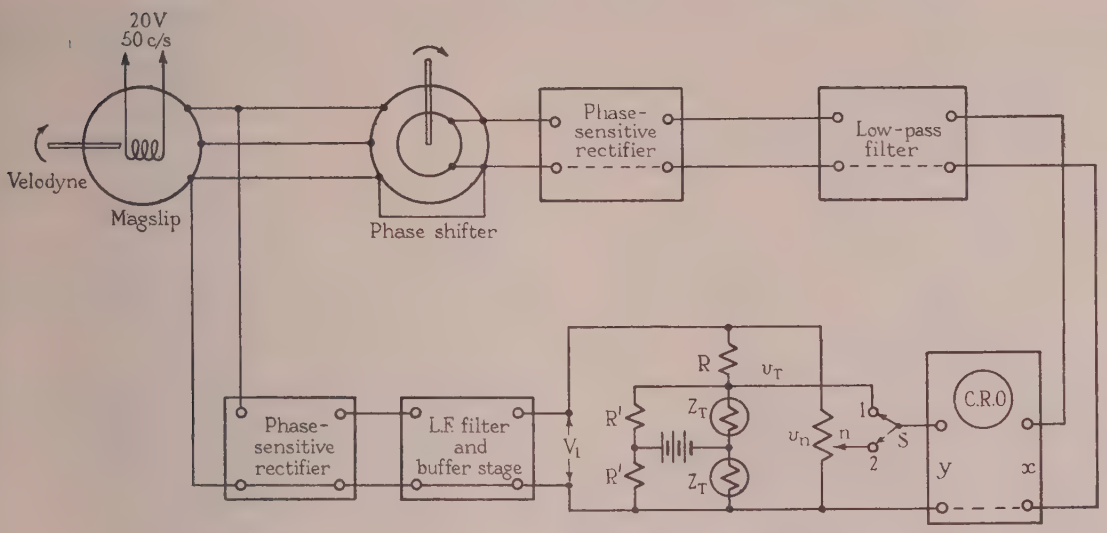


Fig. 8.—Circuit for measuring thermistor impedance.

(6.1) Interpretation of the Results

The value of the thermistor equivalent impedance may now be derived from the relation

$$v_T = v_n \varepsilon^{j\phi} \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

from which

$$\frac{2Z_T R_1}{R'R + Z_T(R + 2R')} = n\varepsilon^{j\phi} \quad . \quad . \quad . \quad (23)$$

Inverting and rearranging, we have

$$\frac{1}{Z_T} = \frac{2}{nR}\varepsilon^{-j\phi} - \frac{R + 2R'}{R'R} \quad . \quad . \quad . \quad . \quad (24)$$

The locus of $1/Z_T$ is now obtained by plotting $(2/nR)\varepsilon^{-j\phi}$ and moving the origin to $(R + 2R')/R'R$. The locus of Z_T is then given by the invert.

In practice it is convenient to make R and R' very much greater than the thermistor impedance, so that n is very small and, since

$$\frac{2}{nR} \gg \frac{R + 2R'}{R'R}$$

therefore

$$Z_T \simeq \frac{nR}{2}\varepsilon^{j\phi} \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

The method may be simplified because the voltage v_n is, in fact, independent of frequency; hence a phase balance with the switch in position 2 need only be obtained for one particular frequency. It is convenient to set the scale so that this balance occurs at zero; the phase of the voltage v_T is then read directly. Furthermore, the amplitude of the vertical deflection of the oscillograph, which is proportional to n, may be calibrated, enabling the amplitude of v_T to be measured directly. These calibrations are checked at intervals during the test, but the fact that the number of readings which need be taken is virtually halved is a great advantage, especially at very low frequencies, where the procedure is laborious.

(6.2) Test Results

Some impedance loci of a bead-type thermistor A1522/100 are plotted in Fig. 9. These loci depend on the value of the polarizing current, as expected. A locus of the variation of impedance with the polarizing current, for several constant frequencies, is given in Fig. 10. The curves show that the phase angle of the thermistor impedance is approximately independent of the polarizing current.

The phase angle of Z_T, as given in eqn. (9), will be independent of the polarizing current provided that R_0/R_∞ remains constant, since τ $(= S/K)$ is fixed.

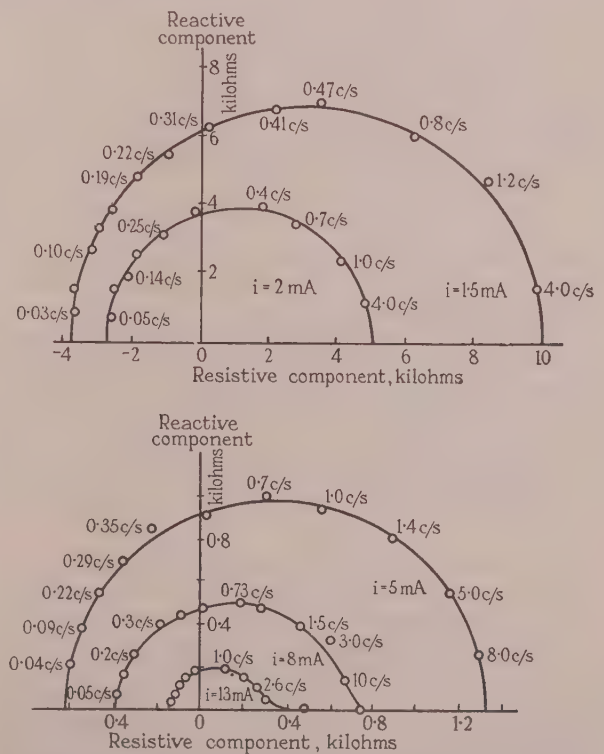


Fig. 9.—Impedance loci of thermistor A1522/100 for various polarizing currents.

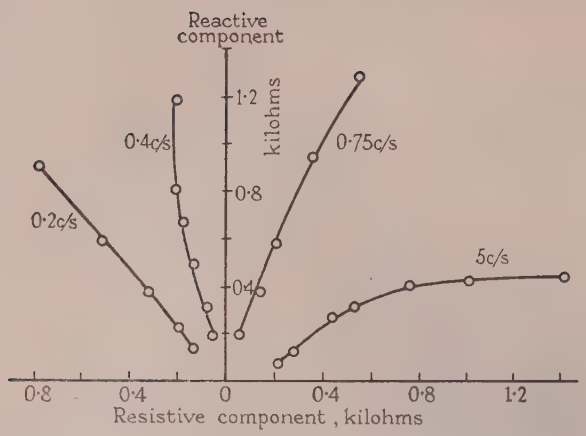


Fig. 10.—Loci of thermistor impedance at constant frequency and varying polarizing current.

If the impedance of a thermistor with a given polarizing current is Z_T, the impedance with any other polarizing current may be expressed approximately as kZ_T, if k, a real number, does not exceed 3.

(7) DISTORTION INTRODUCED BY THE THERMISTOR

(7.1) Harmonic Distortion of the Waveform

In the push-pull arrangement used for polarizing the thermistors in Fig. 8, any even harmonic distortion of the voltage waveform cancels out, since a rise in the net current in one thermistor is accompanied by a fall in the current in the other. An examination of the voltage across one of the thermistors shows that the distortion is negligible provided that the peak signal current is less than a quarter of the polarizing current.

At the higher frequencies, i.e. frequencies at which the impedance of the thermistor may be approximately expressed as R_∞, the distortion is zero even for large inputs. This is to be expected physically, since the harmonic-distortion voltages in eqn. (13) are very small when the frequency is high.

The magnitude of the thermistor impedance, and therefore the voltage across it, due to a constant-amplitude current decreases with decreasing frequency. The harmonic-distortion voltages increase, however, so that the percentage distortion is greatest at very low frequencies.

The extent of the harmonic distortion at low frequencies in any given thermistor may be seen from the static characteristics, since at low frequencies the voltage at any instant is given by the ordinate of the characteristic curve. The distortion will be small for small signal currents, provided that the polarizing current is such that the operating point is not near the hump in the characteristic. If the distortion is small at low frequencies, it may be concluded that it is small at all other frequencies.

(7.2) The Non-Linear Impedance of a Thermistor

At very high frequencies the resistance of the thermistor does not change with individual cyclic changes of current, but assumes a steady value R_∞. There can therefore be no distortion of the waveform. R_∞ is the resistance of the thermistor, and will depend on the total power dissipated. For a given polarizing current and small alternating currents, R_∞ will be approximately constant, but when the power supplied by the alternating current is comparable with that due to the polarizing current, R'_∞ will depend appreciably on the value of this alternating current. An exact proportionality will not then exist between the amplitude of the voltage and the amplitude of the current, as is shown by the curves in Fig. 11, in which the peak alternating voltage

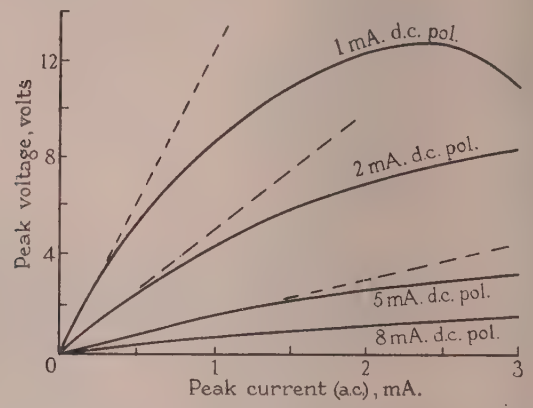


Fig. 11.—Relation between peak alternating voltage and peak alternating current for 50c/s.

$\frac{1}{2}(V_{max} - V_{min})$ is plotted against the peak alternating current, with the thermistor polarized with various steady currents.

The reason for the curvature of these graphs may be seen in the thermistor static characteristic. Assume that a thermistor is polarized with a current represented by ON, in Fig. 12, and that

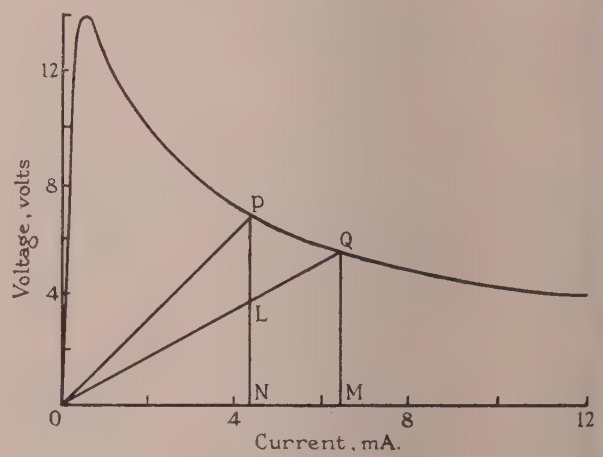


Fig. 12.—Effect of a large-amplitude signal on the thermistor resistance.

an alternating current of very small amplitude and high frequency is superimposed. Then the thermistor impedance to this current is given by the slope of OP. If the amplitude of the alternating current is now increased, the operating point will move from P to Q, where OM is the r.m.s. value of the total current. The thermistor impedance to the alternating current is still resistive, but is now equal to the slope of OQ. It should be noted that this lack of proportionality between the peak alternating voltage and the peak alternating current does not give rise to harmonic distortion, as in the case of valve characteristics.

This movement of the effective operating point takes place not only with applied currents of high frequency, but also at low frequencies. In that case, however, the resistance of the thermistor is negative, and an increase in the alternating current

will bring about a fall in the net power, and it follows that the effective operating point will move towards the ordinate axis. At the frequency at which the thermistor impedance is a pure reactance, no power is supplied by the alternating current, and the operating point will be unaffected by small increases in the amplitude of the signal current.

(7.3) Effect of a Large Signal Current on the D.C. Standing Voltage

The movement of the operating point from P to Q for large signals will result in a change in the direct standing voltage across the thermistor, equal to PL. This is a serious disadvantage, because in low-frequency circuits the conventional resistance-capacitance a.c.-d.c. separating circuits are not practical; the standing voltages must be "backed off" against a reference voltage, and in such a circuit any change in the standing voltage will result in a superfluous d.c. component in the output of the circuit. This difficulty is overcome by employing a push-pull arrangement for polarizing the thermistors, as shown in Fig. 8. Any change in the voltage in one thermistor is then balanced by a similar change in the other thermistor, but in the reverse direction with respect to the output.

(8) STABILITY OF THERMISTOR CIRCUITS

Consider the circuit shown in Fig. 2, in which a thermistor is connected to a source of voltage in series with a resistor R. If the thermistor impedance to incremental changes is given by Z_T [see eqn. (19)] the total impedance in the circuit is

$$Z = R + Z_T \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

$$= \frac{(R - R_0) + D\tau(R + R_\infty)}{1 + D\tau} \quad . \quad . \quad . \quad (27)$$

This will be zero when

$$D = \frac{(R_0 - R)}{\tau(R + R_\infty)} \quad . \quad . \quad . \quad . \quad . \quad (28)$$

If R_0 is greater than R this will be positive, showing that the circuit is unstable.

Since V and R may be regarded as the Thévenin equivalent circuit of a general resistive network, it may be concluded that a general resistive network containing one thermistor will be stable provided that the resistance presented to the thermistor is greater than that corresponding to the slope of its static characteristic at the operating point. It should be noted that the stability of the network is independent of the e.m.f.'s in the circuit except to the extent that they determine the operating point. It is convenient to omit the polarizing e.m.f.'s from the theoretical circuits.

The above criterion of stability is readily tested by using the circuit shown in Fig. 13. The voltage is adjusted so that no current flows in the variable resistor R; R may then be varied without disturbing the polarizing current. As R is reduced the effective load line (see Fig. 1) rotates anticlockwise about the operating point P. At a polarizing current of 2mA the slope of the thermistor characteristic is approximately 2 kilohms. When R is made equal to or less than this value the current in the thermistor drifts to a very small value.

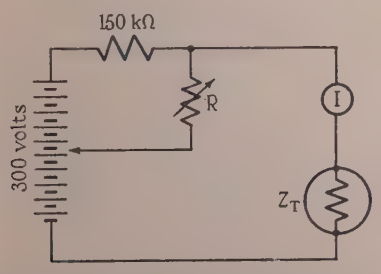


Fig. 13.—Circuit for testing the stability criterion $R > R_0$.

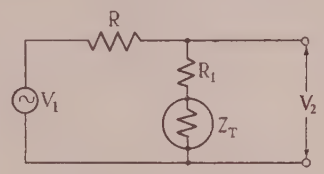


Fig. 14A.—Phase-advance circuit.

(9) A PHASE-ADVANCE NETWORK

Fig. 14A shows the circuit of a differentiating or lead network employing a polarized thermistor; the means of supplying the polarizing current are not shown. The equivalent circuit is

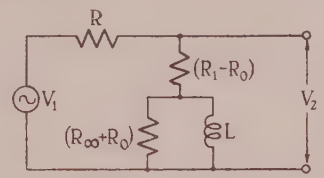


Fig. 14B.—Equivalent circuit of the phase-advance network.

shown in Fig. 14B, and it should be noted that R_1 is greater than R_0. The voltage ratio is given by

$$G' = \frac{V_2}{V_1} = \frac{R_1 + Z_T}{R + R_1 + Z_T}$$

$$= \frac{(R_1 - R_0) + D\tau(R_1 + R_\infty)}{(R + R_1 - R_0) + D\tau(R + R_1 + R_\infty)} \quad . \quad (29)$$

the locus of which is a semicircle, as shown in Fig. 14C.

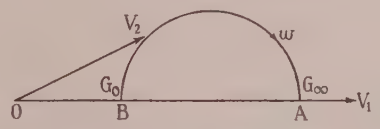


Fig. 14C.—Gain locus of the phase-advance network.

This locus is similar to that of the conventional phase-advance network shown in Fig. 15. This RC circuit is not practical at very low frequencies, as a simple example will suffice to show. For a given ratio of high- to low-frequency gain ratio OA/OB, say 5, the time-constant is $R'C/5$. If the time-constant is to be 10sec then the product $R'C$ must equal 50. The large

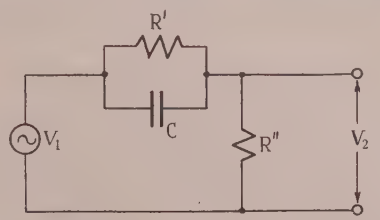


Fig. 15.—Resistance-capacitance phase-advance circuit.

low-leakage capacitors which this would require tend to be inconveniently bulky and expensive.

The thermistor network does not suffer from these disadtages, however, since a suitable choice of series and shunt resistors enables the time-constant to be adjusted to any desired value. For example, the gain of the circuit shown in Fig. 16 is

$$G = \frac{R_1 + \dfrac{R_3(R_2 + Z_T)}{R_3 + R_2 + Z_T}}{R + R_1 + \dfrac{R_3(R_2 + Z_T)}{R_3 + R_2 + Z_T}}$$

$$= \frac{(R_Y + Z_T)(R_3 + R_1)}{(R_X + Z_T)(R_3 + R_1 + R)} \quad . \quad . \quad . \quad (30)$$

where $$R_X = R_2 + \frac{R_3(R + R_1)}{R_3 + R_1 + R} \quad . \quad . \quad . \quad . \quad (31)$$

which is the resistance presented to the thermistor, and

$$R_Y = R_2 + \frac{R_1 R_3}{R_1 + R_3} \quad . \quad . \quad . \quad . \quad (32)$$

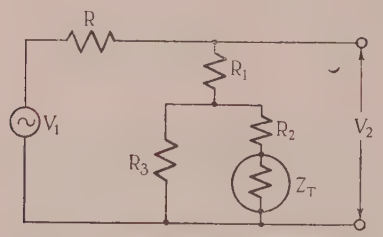


Fig. 16.—General phase-advance network.

This is the resistance presented to the thermistor with the output short-circuited.

The time-constant of the circuit, τ_G, is therefore given by

$$\tau_G = \tau \frac{R_X + R_\infty}{R_X - R_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

It has been shown that if the circuit is to be stable R_X must be greater than R_0. In order that the time-constant shall be large, R_X must be made approximately equal to R_0; thus the larger the time-constant is made the greater is the possibility of the circuit being unstable.

If the circuit had to be used where very large time-constants are involved ($\tau_G > 5\tau$), a thermistor with different thermal properties would be preferable. An increase in the ratio of thermal capacity to dissipation constant, namely S/K, would increase the time-constant. This is always possible, since S is a function of volume while K is a function of surface area.

(9.1) Design of the Phase-Shift Network

Assume that the thermistor impedance is given, that it is required to choose suitable resistors R, R_1, R_2, R_3, so that the ratio of the high-frequency gain to the low-frequency gain G_∞/G_0 is given some specified value y, and that the time-constant τ_G is to be some multiple, x, of τ, the time-constant of the thermistor alone.

Then $$x\tau = \frac{(R_X + R_\infty)}{(R_X - R_0)}\tau \quad . \quad . \quad . \quad . \quad . \quad (34)$$

and $$R_X = \frac{xR_0 + R_\infty}{(x - 1)} \quad . \quad . \quad . \quad . \quad . \quad (35)$$

Hence R_X is fixed, and since

$$y = \frac{G_\infty}{G_0} = \frac{(R_Y + R_\infty)(R_X - R_0)}{(R_X + R_\infty)(R_Y - R_0)} = \frac{(R_Y + R_\infty)}{(R_Y - R_0)}\frac{1}{x} \quad . \quad (36)$$

giving $$R_Y = \frac{R_\infty + xyR_0}{xy - 1} \quad . \quad . \quad . \quad . \quad . \quad (37)$$

R_Y is also fixed.

The resistance R should be made as small as possible so that the attenuation will be small. The minimum value will be determined by the output impedance of the voltage source, and R will therefore first be fixed at a convenient value. Then either R_1, R_2, or R_3 may be given an arbitrary value, say zero. Whichever is chosen, the others will be fixed by eqns. (35) and (37). By way of example, assume that a simple phase-advance network is to be designed, using a thermistor so polarized that $R_\infty = 3$ kilohms, $R_0 = 1$ kilohm, and $\tau = 1$ sec. The ratio of the high-frequency to the low-frequency gain should be 5, and the time-constant 5 sec. The network is to be fed from a source having an impedance of 10 kilohms, and is to feed into an impedance of 2 kilohms. The source impedance may be made equal to R (10 kilohms) and R_1 may be zero; then the impedance of the load will be part of R_3.

From eqns. (35) and (37), $R_X = 2$ kilohms and $R_Y = 7/6$ kilohms; therefore $R_2 = 2$ kilohms and $R_3 = 10/11$ kilohms. From eqn. (30) the high- and low-frequency gains are $G_\infty = 0{\cdot}07$ and $G_0 = 0{\cdot}014$.

An RC circuit feeding into an impedance of 2 kilohms with a time-constant of 5 sec would be quite impracticable, and a cathode-follower buffer stage would be required. In the thermistor circuit the attenuation appears to be large, but is not so in fact, because with the phase-shift circuit removed there is an attenuation factor of 6 due to the source impedance.

(10) A SIMPLE INTEGRATING CIRCUIT

The gain of the general phase-retarding network shown in Fig. 17 is

$$\frac{V_2}{V_1} = 1 - \frac{V_3}{V_1} = 1 - G \quad . \quad . \quad . \quad . \quad . \quad (38)$$

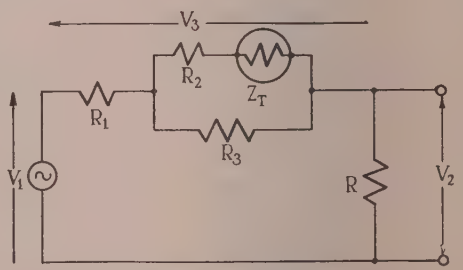


Fig. 17.—General form of the phase-retarding network.

because $V_3/V_1 = G$ is the gain of the differentiating network shown in Fig. 16.

The time-constant of the integrating network is therefore the same as that of the differentiating network [eqn. (33)]. Therefore eqn. (35) holds, i.e.

$$R_X = R_2 + \frac{R_3(R_1 + R)}{R + R_1 + R_3} = \frac{xR_0 + R_\infty}{(x - 1)} \quad . \quad . \quad (35)$$

The ratio, Z, of the high- to low-frequency gains is

$$z = \frac{1 - G_\infty}{1 - G_0} = \frac{1 - \dfrac{(R_Y + R_\infty)(R_3 + R_1)}{(R_X + R_\infty)(R_3 + R_1 + R)}}{1 - \dfrac{(R_Y - R_0)(R_3 + R_1)}{(R_X - R_0)(R_3 + R_1 + R)}}$$

$$= \frac{1}{x} \frac{R_X(R_3 + R_1 + R) - R_Y(R_3 + R_1) + R_\infty R}{R_X(R_3 + R_1 + R) - R_Y(R_3 + R_1) - R_0 R}$$

$$= \frac{1}{x} \frac{(R_2 + R_3 + R_\infty)}{(R_2 + R_3 - R_0)} \quad . \quad (39)$$

from which $$(R_2 + R_3) = \frac{zxR_0 + R_\infty}{zx - 1} \quad . \quad . \quad . \quad . \quad (40)$$

Eqns. (35) and (40) may be used to design the integrating networks in the same way as eqns. (35) and (37) were used for the differentiating network.

(10.1) A Phase-Inverting Circuit

When the resistor R_1 in the circuit shown in Fig. 14A is less in magnitude than the thermistor low-frequency resistance R_0 the circuit will invert the sense of a steady signal. For example, consider the circuit shown in Fig. 18A in which R_1 is zero.

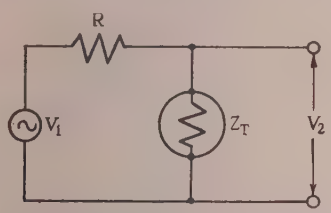


Fig. 18A.—Phase-inverting network.

The gain of the circuit is

$$G = \frac{Z_T}{R + Z_T} \quad . \quad . \quad . \quad . \quad . \quad . \quad (41)$$

and at very low frequencies

$$G_0 = \frac{-R_0}{R - R_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad (42)$$

where R is greater than R_0, in order that the circuit may be stable. Therefore, G_0 is negative. The locus of the gain is shown in Fig. 18B.

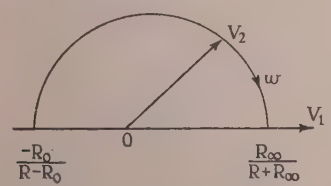


Fig. 18B.—Gain locus of the phase-inverting network.

The output voltage leads the input by an angle which varies from 180° to 0°. Because the sense of a steady voltage is reversed, it is convenient to regard the circuit as a phase-retarding plus a phase-inverting network. In closed-loop control systems the phase reversal of the output voltage must be corrected by reversing the connections in order that the feedback shall remain negative to steady signals.

(11) PRACTICAL FORMS OF THE PHASE-SHIFT CIRCUITS

The method used to polarize the thermistor in a practical circuit will depend on the particular application; some examples

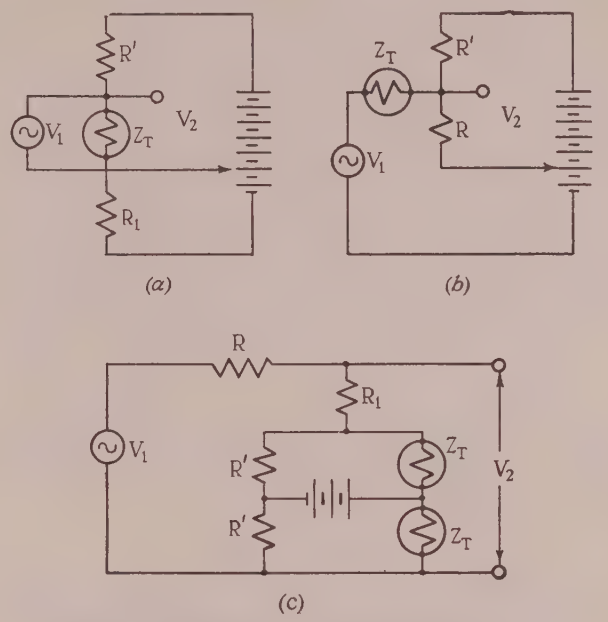


Fig. 19.—Practical form of the phase-shift circuits.

are given in Fig. 19. In all these circuits the resistance R' effectively shunts the thermistor, and its effect is to modify the circuit constants. If this is detrimental to the successful operation of the circuit R' must be large enough to ensure that the effect is negligible.

In Figs. 19(*a*) and 19(*b*), which illustrate a differentiating and an integrating network respectively, the tapping on the battery is adjusted to give zero direct voltage across the output. In Fig. 19(*c*) two thermistors are used with a floating battery. This circuit has the advantage that when the current in one thermistor increases the current in the other decreases; they are, so to speak, operating in push-pull, and distortion due to even harmonics will cancel out.

The circuit employed to measure the thermistor impedance loci (Fig. 8) may be used to plot the gain loci of the phase-shift circuits, the circuit shown in Fig. 20 replacing the thermistor

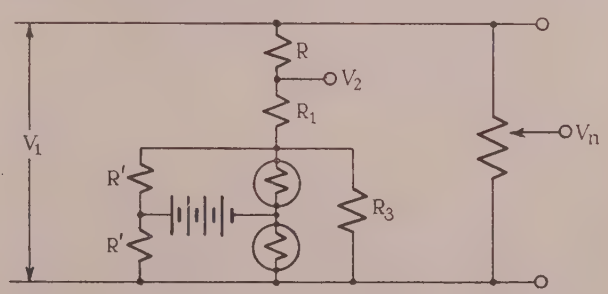


Fig. 20.—Circuit for measuring the gain of the phase-shift circuits.

circuit. The voltages V_1 and V_2 are compared by means of the technique described in Section 6. The results obtained agree closely with the theory.

(12) IMPEDANCE OF A THERMISTOR TO AMPLITUDE-MODULATED SIGNALS

So far, thermistors have been considered as polarized by a direct current and the signal as a variation in this current If, however, instead of d.c. polarization and signal, the thermistor is polarized with alternating current (50c/s), and a modulated signal (of the same frequency and in phase with the polarizing current) is applied, the impedance of the thermistor to the modulation envelope is the same as the impedance discussed in Section 5. This is because the thermistor impedance is a result of the variation in the thermistor resistance, and, provided that the carrier frequency is sufficiently high, the resistance varies in the same way whether alternating or direct currents flow in the thermistor. This is shown by the fact that the static characteristics for 50-c/s currents are approximately the same as those obtained with direct currents. The discussion presented in Section 5 may be applied to the a.c.-polarized thermistor, and the impedance of thermistors to the modulation of an amplitude-modulated signal may therefore be expressed in the form of eqn. (19).

The circuit design for use in d.c. circuits may be used with equal success in circuits where amplitude-modulated signals are employed. The thermistors must of course be polarized with a steady alternating current of the same frequency as the carrier and in phase with it.[10]

(13) CONCLUSION

It has been shown that thermistors may be used to obtain phase shifts between very-low-frequency voltages and currents. Thermistor circuits have several advantages over the conventional *RCL* circuits, namely that very large time-constants are easily realizable; the circuits may be used in a.c. or d.c. systems; and they have properties arising from the fact that the thermistor acts as a negative resistance to very-low-frequency signals.

The use of thermistors has been illustrated by designing a simple phase-shift network. The large time-constant associated with these networks makes them admirably suitable for stabilizing many automatic control systems where the natural frequencies are low. Some examples of such applications are in the steering of ships and of aircraft, where the small size of the thermistor is a great advantage; in process controls; and in circuits where thermistors are already being used as measuring elements, such as in the temperature control of ovens or the stabilization of supply voltages.

In addition to applications in the stabilization of control systems there are many other uses for the thermistor. Circuits for the investigation of recurrent low-frequency natural phenomena, e.g. the study of ocean waves, the measurement of heartbeat and breathing frequencies, and certain geophysical applications, are examples. The author has used thermistors as the frequency-determining elements in an ultra-low-frequency oscillator (0·2c/s–0·02c/s), the basic circuit of which consists solely of resistors and valves and uses no capacitors or inductors.[10]

It is claimed[1] that thermistors with much smaller time-constants than those at present available can be manufactured. With such thermistors the applications of thermal time-constant circuits would not be limited to the ultra-low-frequency range (less than 1c/s) but could be used in the stabilization of most control systems. This would be of great advantage in systems where a.c. data transmission is employed.

(14) ACKNOWLEDGMENT

The work described in the paper was carried out at the Electrical Engineering Department of the University College of North Wales, Bangor, with the aid of a D.S.I.R. research grant. The author wishes particularly to acknowledge his indebtedness to the director of his studies, Dr. David Morris, who was responsible for the initiation of the work, for his advice throughout the preparation of the paper.

(15) REFERENCES

(1) Becker, J. A., Green, C. B., and Pearson, G. L.: "Properties and Uses of Thermistors—Thermally Sensitive Resistors," *Bell System Technical Journal*, 1947, **26**, p. 170; and *Electrical Engineering*, 1946, **19**, p. 711.

(2) Patchett, G. N.: "Precision A.C. Voltage Stabilisers," *Electronic Engineering*, 1950, **22**, pp. 371 and 424.

(3) Stone, J. E.: "An Ultra-Low Frequency Oscillator," *Electronics*, 1950, **23**, p. 94.

(4) Bollman, J. H., and Kreer, J. G.: "The Application of Thermistors to Control Networks," *Proceedings of the Institute of Radio Engineers*, 1950, **38**, p. 20.

(5) Smith, O. J. M.: "Thermistors," *Review of Scientific Instruments*, 1950, **21**, p. 344.

(6) Cooper, W. H., and Seymour, R. A.: "Temperature-dependent Resistors; Uses as Electric Circuit Elements," *Wireless Engineer*, 1947, **24**, p. 298.

(7) Medlock, R. S.: "Automatic Control in the Chemical Industry" (Discussion), *Journal I.E.E.*, 1947, **94**, Part IIA, p. 98.

(8) Tustin, A.: "The Possibility of using the Thermal Time Constant of a Temperature Sensitive Resistor to obtain Phase Advance of an Electrical Signal," C.S. Memo 68, 1943 (Metropolitan-Vickers, Sheffield).

(9) Candy, C. J. N.: "Electrical Networks incorporating Thermal Time Constants." Dissertation for B.Sc. (Honours) Examination, Bangor, 1951.

(10) Candy, C. J. N.: "The Investigation and Development of Systems involving Large Time Constants." Ph.D. Thesis, Bangor, 1954.

(11) Morris, D.: "A Phase-Calibrated Variable-Frequency Supply for the Testing of Servo Mechanisms," *Proceedings I.E.E.*, Paper No. 935M, February, 1950 (**97**, Part II, p. 37).

(12) Dickinson, C. J.: "Thermistor Continuous Temperature Control for Biological Research," *Electronic Engineering*, 1949, **21**, p. 408.

(16) APPENDICES

(16.1) Effect on Thermistor Impedance of Variations in Supply Voltage and of Changes in Ambient Temperature

So far, only changes in the thermistor resistance due to the passage of a signal current through the polarized thermistor have been considered. The resistance is also dependent on the ambient temperature and the polarizing current. Any variation in the ambient temperature or in the supply voltage will therefore affect the impedance of the thermistor to the signal current.

A variation of the supply voltage will change the quiescent conditions of the thermistor to those given by a different point on the static characteristic. The high- and low-frequency resistances R_∞ and R_0 will then be defined by the new operating point, but the time-constant τ will remain unchanged, as shown by eqn. (18).

The effect of temperature changes can be seen from the static characteristics in Fig. 21. These are plotted for various ambient temperatures. The values of R_∞ and R_0 may be determined from these curves for any given polarizing current. It is seen that in general a change of temperature will result in a change in R_∞ and R_0, but the time-constant will remain unaltered.

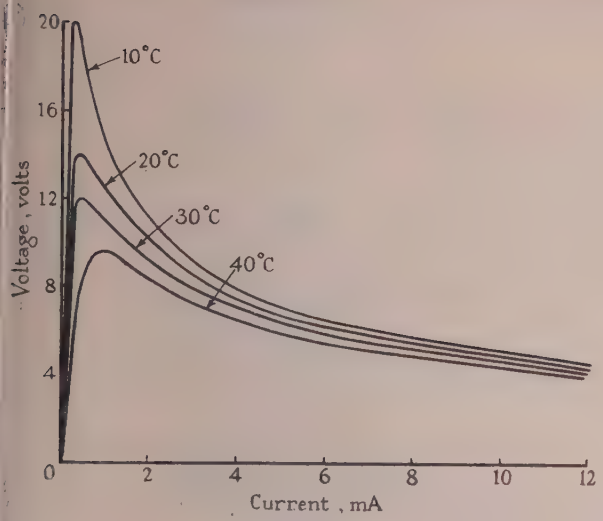


Fig. 21.—Static characteristics of thermistor A1522/100 at various ambient temperatures.

(16.2) Effect of Changes in the Thermistor Impedance on the Properties of Thermistor Networks

Consider the circuit in Fig. 2, in which the polarized thermistor is connected in series with a resistor. The total resistance to high-frequency variations is

$$R + R_\infty = Z_\infty$$

and the percentage change in Z_∞ due to a change of $k\%$ in R_∞ will be $\frac{kR_\infty}{R + R_\infty}\%$, which is small if k is small. At very low frequencies the total impedance is

$$R - R_0 = Z_0$$

and the percentage change in Z_0 due to a change of $k\%$ in R_0 will be $\frac{kR_0}{R - R_0}\%$ and this will be much greater than k if $R \simeq R_0$.

We conclude that in circuits where the effective resistance presented to the thermistor is comparable with the magnitude of the thermistor low-frequency resistance R_0, even small changes in the ambient temperature, or in the supply voltage, may have a pronounced effect on the properties of the circuit. The circuit will become unstable if R_0 increases sufficiently to make $R_0 > R$. Circuits in which $R \simeq R_0$ must be fed from stabilized power sources, and the thermistors should be enclosed in a constant-temperature container.

(16.3) Changes in the D.C. Balance Conditions

In general, a change in the polarizing current or in the temperature will result in a change in the standing voltage across the thermistor (see Section 7.3). This will result in a spurious output voltage unless a push-pull arrangement is employed for polarizing the thermistors.

(16.4) Use of Thermistors as Measuring Elements

Thermistors may be used as thermometers[12] or as a means of measuring changes in a supply voltage.[2] A typical circuit is

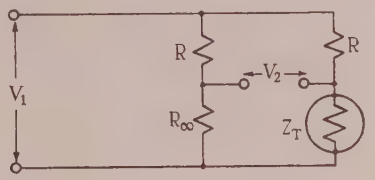


Fig. 22.—A voltage-sensitive bridge.

shown in Fig. 22. A change in the voltage V_1 will produce an out-of-balance voltage given by

$$V_2 = V_1\left(\frac{R_\infty}{R + R_\infty} - \frac{Z_T}{Z_T + R}\right)$$

from which $$\frac{V_2}{V_1} = \frac{(R_\infty + R_0)R}{(R + R_\infty)(R - R_0) + D\tau(R_\infty + R)}$$

The output therefore lags behind the input; the time-constant of the circuit is $\tau(R_\infty + R)/(R - R_0)$ and the steady-state gain at low frequencies is $R(R_\infty + R_0)/(R + R_\infty)(R - R_0)$.

These results show that the sensitivity is increased by reducing R (or by shunting the thermistor, which increases the effective value of R_0). The expression for the time-constant shows that the lag increases as R decreases; this may be a serious disadvantage when the circuit is part of a closed system in which an increase in the time-constant may bring about instability.

[The discussion on the above paper will be found on page 415.]

621.376.2 : 621.3.012.1

Paper No. 1772 M
Jan. 1955

A VECTOR METHOD FOR AMPLITUDE-MODULATED SIGNALS

By C. J. N. CANDY, Ph.D., Graduate.

(*The paper was first received* 15*th June, and in revised form* 28*th September*, 1954. *It was published in January*, 1955, *and was read before a Joint Meeting of the* MEASUREMENT AND CONTROL SECTION *and the* RADIO AND TELECOMMUNICATION SECTION 29*th November*, 1955.)

SUMMARY

The paper describes a vector notation for amplitude-modulated alternating quantities. This notation enables vector expressions to be derived for the impedances, admittances and transfer ratios of electrical circuits, and these expressions in turn operate directly on the vector for the amplitude-modulated voltages and currents to give the vectors for the dependent variables in the circuit. The results obtained by this vector analysis are in a convenient form for the direct calculation of the output of a demodulator, for any general input to the circuit. Examples are given to show that when the modulation frequency is comparable with the carrier frequency the modulation (or signal) may be appreciably modified by impedances in the carrier channel. The paper includes a description of a complementary operational method, which enables the response of circuits to non-sinusoidal amplitude-modulated signals to be determined.

LIST OF PRINCIPAL SYMBOLS

ω_m = Angular frequency of the modulation.
ω_c = Angular frequency of the carrier.
$\mathscr{R}\boldsymbol{Y}$ = Real part of a vector $\boldsymbol{Y}$.
j, j_m, j_c = Complex operators.
$\mathrm{D} = \dfrac{d}{dt}$.
$\boldsymbol{G} = A + j_cB$ = Transfer impedance.
A, B = Functions of j_m but independent of j_c.
ϕ_R = Phase angle by which a reference voltage leads the carrier.
$\mathrm{D_m}, \mathrm{D_c}$ = Bifid operators.

(1) INTRODUCTION

The calculation of the response of a circuit to an amplitude-modulated signal by the normal methods of circuit analysis (e.g. Heaviside's operational calculus or Laplace transforms) tends to be laborious. An approximate method has been used for a limited class of circuits, as in References 1, 2, and 3, for example. An exact method which may be applied to any type of circuit has been presented in Reference 4; in this the amplitude-modulated quantity is expressed as the sum of two sidebands, and the response of the network to these is found separately by the normal methods of harmonic analysis. The output of the network is then determined by adding these two responses.

The vector method described in the paper enables a general expression for the impedance of networks to be defined, and by this method the impedance of a composite network can be derived from the separate impedances of the constituent networks; this is not easily accomplished by the methods cited above. A further advantage is that the output of a demodulator may be readily obtained in a form which clearly indicates the effect on the demodulated signal of the phase angle between the reference voltage and the signal carrier.

The vector notation is used to represent suppressed-carrier amplitude-modulated quantities as encountered in servo-mechanisms. The results can, however, be applied to the more general modulation encountered in communication circuits, since these may be represented by the sum of a carrier amplitude-modulated quantity and an unmodulated carrier, thus

$$(1 + m\cos\omega_m t)\cos\omega t \equiv m\cos\omega_m t\cos\omega_c t + \cos\omega_c t \quad . \quad (1)$$

The response of the circuit to these signals may be determined separately.

(2) A VECTOR NOTATION FOR AMPLITUDE-MODULATED ALTERNATING QUANTITIES

A simple harmonic quantity may be denoted by the real component of a rotating vector, thus

$$\cos\omega t \equiv \mathscr{R}[\varepsilon^{j\omega t}] \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where j is an operator which advances the phase of the vector by an angle of $\pi/2$ ($j = \varepsilon^{j\pi/2}$) and j^2 is numerically equal to (-1). It is not correct to identify j with $\sqrt{(-1)}$ since j is by definition an operator, as is pointed out in Reference 7. It is permissible to define any number of operators similar to j each representing rotations in different planes. If j_m and j_c are two such operators, we may represent sinusoidal amplitude-modulated quantities by the real part of the product of two vectors rotating in planes at right angles to each other, thus

$$\cos\omega_m t\cos\omega_c t \equiv \mathscr{R}[\varepsilon^{j_m\omega_m t}\varepsilon^{j_c\omega_c t}] \equiv \mathscr{R}[\varepsilon^{(j_m\omega_m + j_c\omega_c)t}] \quad . \quad (3)$$

where the real part is taken to be the sum of all components of the vector that can be expressed as real functions, i.e. independent of both j_m and j_c.

Multiplication of a vector by j_m or j_c advances the phase of the modulation and the carrier respectively by $\pi/2$. j_m^2 and j_c^2 are numerically equal to -1; j_mj_c represents a phase advance of both the modulation and the carrier, as is shown by the following equations. Let

$$\begin{aligned}\boldsymbol{Y} = \varepsilon^{(j_m\omega_m t + j_c\omega_c t)} &= (\cos\omega_m t + j_m\sin\omega_m t)(\cos\omega_c t + j_c\sin\omega_c t)\\ &= \cos\omega_m t\cos\omega_c t + j_m\sin\omega_m t\cos\omega_c t\\ &\quad + j_c\cos\omega_m t\sin\omega_c t + j_mj_c\sin\omega_m t\sin\omega_c t \quad . \quad (4)\end{aligned}$$

Then

$$\begin{aligned}\mathscr{R}\boldsymbol{Y} &= \cos\omega_m t\cos\omega_c t\\ \mathscr{R}(j_m\boldsymbol{Y}) &= -\sin\omega_m t\cos\omega_c t = \cos(\omega_m t + \pi/2)\cos\omega_c t\\ \mathscr{R}(j_c\boldsymbol{Y}) &= -\cos\omega_m t\sin\omega_c t = \cos\omega_m t\cos(\omega_c t + \pi/2)\\ \mathscr{R}(j_mj_c\boldsymbol{Y}) &= \sin\omega_m t\sin\omega_c t = \cos(\omega_m t + \pi/2)\cos(\omega_c t + \pi/2)\\ \mathscr{R}(j_m^2\boldsymbol{Y}) &= -\cos\omega_m t\cos\omega_c t = -\mathscr{R}\boldsymbol{Y}\end{aligned}$$

The vector $\boldsymbol{Y}$ may be represented geometrically, but since it consists of four components each of which must be depicted in quadrature with the other three the vector cannot be shown in a three-dimensional diagram. The vector is shown as a series of two-dimensional projections in Fig. 1.

Fig. 1(*a*) shows a projection of the vector in the plane of j_c and the real axis. We have

$$|\mathrm{OA}| = \cos\omega_m t\cos\omega_c t \text{ and } |\mathrm{AB}| = \cos\omega_m t\sin\omega_c t$$

so that $|\mathrm{OB}| = \cos\omega_m t$ and angle $\mathrm{AOB} = \omega_c t$

Dr. Candy is with Messrs. S. Smith and Sons. Ltd., and was formerly at the University College of North Wales, Bangor.

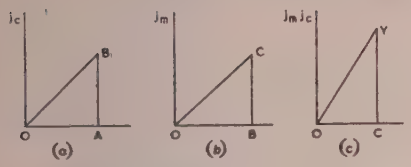


Fig. 1.—Projections of the vector Y.

Fig. 1(b) shows the projection in the plane of OB [from Fig. 1(a)] and the j_m axis. Since

$$|\mathrm{OB}| = \cos \omega_m t, \text{ and } |\mathrm{BC}| = \sin \omega_m t \cos \omega_c t$$

therefore $$|\mathrm{OC}| = \sqrt{[(1 - \sin^2 \omega_m t \sin^2 \omega_c t)]}$$

Fig. 1(c) shows the projection in the plane of OC and the $j_m j_c$ axis. It will be seen that

$$|\mathrm{CY}| = \sin \omega_m t \sin \omega_c t$$

so that $$|\mathrm{OY}| = 1$$

(3) SOME EXAMPLES OF THE USE OF THE VECTOR NOTATION

(3.1) Impedance of a Pure Inductance to Amplitude-Modulated Alternating Signals

The voltage across a pure inductance, L, carrying a current

$$i = \hat{I} \cos \omega_m t \cos \omega_c t = \mathscr{R}[\hat{I} \varepsilon^{(j_m\omega_m + j_c\omega_c)t}] \quad . \quad . \quad (5)$$

will be given, in vector notation, by

$$V = L\frac{dI}{dt} = (j_m\omega_m + j_c\omega_c)LI \quad . \quad . \quad . \quad . \quad (6)$$

(A discussion of the validity of this process of operating on the vector is given in Section 11.)

Similarly the current flowing in the inductance when the voltage is given vectorially by $V = \hat{V}\varepsilon^{(j_m\omega_m + j_c\omega_c)t}$ is

$$I = \frac{1}{L}\int V dt = \frac{V}{(j_m\omega_m + j_c\omega_c)L} \quad . \quad . \quad . \quad (7)$$

Thus the inductance may be regarded as having a vector impedance $(j_m\omega_m + j_c\omega_c)L$ to sinusoidal amplitude-modulated quantities; similarly a condenser has an impedance $1/(j_m\omega_m + j_c\omega_c)C$. These expressions become the conventional impedance $j\omega L$ and $1/j\omega C$ when either ω_m or ω_c tends to zero and the amplitude-modulated quantities become sine waves.

(3.2) Instantaneous Currents and Voltages

The instantaneous voltage across the inductance when the current is defined by eqn. (5) will be the real part of eqn. (6), or

$$v = -\hat{I}L(\omega_m \sin \omega_m t \cos \omega_c t + \omega_c \cos \omega_m t \sin \omega_c t) \quad . \quad (8)$$

and the instantaneous current due to a voltage $\hat{V} \cos \omega_m t \cos \omega_c t$ will be given by the real part of eqn. (7), thus

$$i = \mathscr{R}\left[\frac{\hat{V}\varepsilon^{(j_m\omega_m + j_c\omega_c)t}}{(j_m\omega_m + j_c\omega_c)L}\right] = \mathscr{R}\left[-\frac{\hat{V}(j_m\omega_m - j_c\omega_c)\varepsilon^{(j_m\omega_m + j_c\omega_c)t}}{(\omega_m^2 - \omega_c^2)L}\right]$$

so that

$$i = \hat{V}\frac{(\omega_m \sin \omega_m t \cos \omega_c t - \omega_c \cos \omega_m t \sin \omega_c t)}{(\omega_m^2 - \omega_c^2)L} \quad . \quad (9)$$

In both results [eqns. (8) and (9)] the quantities are expressed as the sum of two amplitude-modulated waves. In general, the currents and voltages in any branch of a linear network due to an applied amplitude-modulated voltage may be expressed as the sum of two amplitude-modulated waves (see Reference 4). It is convenient to arrange these expressions so that the carrier of one wave is in phase with the applied carrier, while the carrier of the other is in quadrature with it. This can always be achieved by rationalizing the vector expression with respect to j_c as in the following example.

(3.3) Example of Current and Voltage Calculation

Let it be required to find the current in a coil having resistance R and inductance L when an amplitude-modulated voltage $v = \hat{V} \cos \omega_m t \cos \omega_c t$ is applied to it. The vector current will be

$$I = \frac{V}{R + \mathrm{D}L}$$

which is of the form

$$I = \frac{KV}{1 + DT} = \frac{KV}{1 + (j_m\omega_m + j_c\omega_c)T} \quad . \quad . \quad (10)$$

so that

$$I = K\left[\frac{(1 + j_m\omega_m T) - j_c\omega_c T}{(1 + j_m\omega_m T)^2 + \omega_c^2 T^2}\right]V = (A + j_c B)V \quad . \quad (11)$$

where $$A = \frac{K(1 + j_m\omega_m T)}{(1 + j_m\omega_m T)^2 + \omega_c^2 T^2}$$

and $$B = \frac{-KT\omega_c}{(1 + j_m\omega_m T)^2 + \omega_c^2 T^2}$$

Then the instantaneous current will be

$$i = i' \cos \omega_c t - i'' \sin \omega_c t \quad . \quad . \quad . \quad . \quad (12)$$

where $$i' = \hat{V}\mathscr{R}(A\varepsilon^{j_m\omega_m t}) \text{ and } i'' = \hat{V}\mathscr{R}(B\varepsilon^{j_m\omega_m t}) \quad . \quad . \quad (13)$$

This result is arrived at by finding the real part of I. This may be determined by first regarding j_m as being real and then determining the real part, namely

$$\hat{V}[A \cos \omega_c t - B \sin \omega_c t]\varepsilon^{j_m\omega_m t} \quad . \quad . \quad . \quad (14)$$

The final result is then obtained by again calculating the real part, this time regarding j_m as being imaginary.

The instantaneous values of i' and i'' are given by

$$i' = K\hat{V}\left\{\frac{(1 + \omega_m^2 T^2)}{[1 + (\omega_m^2 - \omega_c^2)T^2]^2 + 4\omega_m^2 T^2}\right\}^{\frac{1}{2}} \cos\left[\omega_m t + \arctan \omega_m T - \arctan \frac{2\omega_m T}{1 + (\omega_c^2 - \omega_m^2)T^2}\right]$$

$$i'' = \frac{-\omega_c T\hat{V}K}{\left\{[1 + (\omega_m^2 - \omega_c^2)T^2]^2 + 4\omega_m^2 T^2\right\}^{\frac{1}{2}}} \cos\left[\omega_m t - \arctan \frac{2\omega_m T}{1 + (\omega_c^2 - \omega_m^2)T^2}\right]$$

These instantaneous values are given here for completeness; in general, the vector expressions given in eqn. (13) are more useful in circuit analysis.

The vector notation has been used so far to determine the currents and voltages in a given impedance. It may be used in exactly the same way to determine the output of a four-terminal network when an amplitude-modulated voltage is applied to its input terminals. For example, the output of a four-terminal

network having a transmission ratio $K/(1 + DT)$ and input voltage $v_1 = \hat{V} \cos \omega_m t \cos \omega_c t$ is given by

$$V_2 = V_1(A + j_c B)$$

where A and B are defined by eqn. (11).

(4) FUNCTION OF A DETECTOR

An efficient detector may be regarded as a circuit for multiplying the amplitude-modulated signal by a reference voltage of carrier frequency, followed by a low-pass filter which rejects all signals of frequency equal to or greater than the carrier frequency.

Let the reference voltage be $\hat{V}_R \cos(\omega_c t + \phi_R)$ and the signal voltage $\hat{V} \cos \omega_m t \cos \omega_c t$. Then the output will be proportional to the low-frequency component of the product

$$\hat{V}\hat{V}_R \cos \omega_m t \cos \omega_c t \cos(\omega_c t + \phi_R)$$
$$= \frac{\hat{V}\hat{V}_R}{2}[\cos \omega_m t \cos \phi_R + \cos \omega_m t \cos(2\omega_c t + \phi_R)]$$

so that
$$v_2 = \hat{V} \cos \omega_m t \cos \phi_R \quad . \quad . \quad . \quad . \quad (15)$$

provided that $\omega_m \ll \omega_c$.

The constant of proportionality is omitted from the above expression, since any attenuation of the signal may be included as part of the gain of the amplifiers preceding the detector.

If the input voltage were $\hat{V} \cos \omega_m t \sin \omega_c t$ the output would be given by

$$v_2 = -\hat{V} \cos \omega_m t \sin \phi_R \quad . \quad . \quad . \quad . \quad (16)$$

This equation represents the idealized condition, in which it is assumed that the detector obeys the superposition theorem.

In most servo-mechanism applications the signal voltage consists of a suppressed-carrier modulated wave of the form so far discussed. The reference voltage is injected into the detector, as, for example, the bias on the diode circuit of a phase-sensitive rectifier, or as a standing current in the quadrature field of a two-phase induction motor, in which case the reference voltage must be phase-advanced by $\pi/2$ since maximum torque occurs when the carrier is in quadrature with the reference voltage.

In communication circuits it is common practice to have signals of the form given in eqn. (1), i.e. a reference voltage is included in the signal and the phase of this with respect to the carrier is fixed by the circuits in the carrier channel.

(5) TRANSMISSION RATIO OF A TYPICAL CIRCUIT

Consider the response of the network shown in Fig. 2 to an input signal v_1. The transmission ratio of the carrier channel

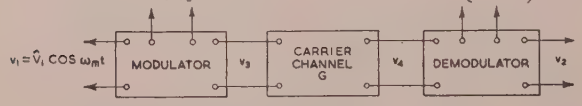


Fig. 2.—An a.c. data transmission system.

(assumed to be linear) is G, which includes the attenuation of the filter circuits following the modulator and the attenuation of the demodulator. G may always be expressed in the general form $A + j_c B$.

The input to the carrier channel will be $v_3 = v_1 \cos \omega_c t$ or, in vector form, $V_3 = V_1 \varepsilon^{j_c \omega_c t}$, and the output of the carrier channel will be given by

$$V_4 = G \times V_3 = \hat{V}_1(A + j_c B)\varepsilon^{(j_m \omega_m + j_c \omega_c)t}$$

Therefore v_4 may be written in the form

$$v_4 = v_4' \cos \omega_c t - v_4'' \sin \omega_c t$$

where $v_4' = \mathscr{R}[\hat{V}_1 A \varepsilon^{j_m \omega_m t}]$ and $v_4'' = \mathscr{R}[\hat{V}_1 B \varepsilon^{j_m \omega_m t}]$

From eqns. (15) and (16) the output of the demodulator will be

$$v_2 = v_4' \cos \phi_R + v_4'' \sin \phi_R$$

This may be written in vector form thus

$$V_2 = (A \cos \phi_R + B \sin \phi_R)V_1 \quad . \quad . \quad . \quad (17)$$

The whole circuit may therefore be regarded as having a vector transmission ratio of

$$Z = A \cos \phi_R + B \sin \phi_R \quad . \quad . \quad . \quad . \quad (18)$$

If the circuit forms part of a closed-loop servo-mechanism, Z will be multiplied by the transmission ratio of the d.c. side (this will include the attentuation of the filter in the output of the demodulator) to give the net loop-gain which is used to determine the stability of the circuit.

For communication circuits a curve showing the variation of $|Z|$ with frequency will give an indication of the distortion introduced into the signal by the reactances in the carrier channel.

The following example will serve to make the method clearer.

Let G be of the form $K/(1 + DT)$, where K is a constant. Then from eqns. (11) and (18) we may write the transfer impedance of the circuit in the form

$$Z = K\frac{(1 + j_m \omega_m T) \cos \phi_R - \omega_c T \sin \phi_R}{(1 + \omega_c^2 T^2) + 2j_m \omega_m T - \omega_m^2 T^2} \quad . \quad . \quad (19)$$

(It is sometimes convenient to omit the suffix m from the j in this expression since this impedance acts only on the modulation.) When ϕ_R has the values 0, arc tan $(-\omega_c T)$, and 90° + arc tan $(-\omega_c T)$ the values of Z will be those given in eqns. (12), (14) and (15) on p. 525 of Reference 4.

Eqn. (19) enables ϕ_R to be chosen so as to give an optimum response. For example, if ϕ_R is made equal to arc tan $(1 - \omega_c^2 T^2)/2\omega_c T$, Z will be approximately equal to $(\omega_c TK)/(1 + \omega_c^2 T^2)$, provided that $\omega_m^2 \ll \omega_c^2$. There are no phase shifts of the signal.

In communication circuits the phase of the reference voltage with respect to the sending-end carrier is fixed by the circuits in the carrier channel. Therefore $\phi_R = -$ arc tan $\omega_c T$, assuming that the carrier and the reference voltage are in phase at the input to the carrier channel, as is the usual practice. Under these conditions,

$$Z = \frac{K}{\sqrt{(1 + \omega_c^2 T^2)}} \frac{1 + \omega_c^2 T^2 + j_m \omega_m T}{1 + \omega_c^2 T^2 + 2j_m \omega_m T - \omega_m^2 T^2}$$

If ω_m is comparable in magnitude with ω_c this circuit will introduce considerable distortion into the signal, unless $\omega_c T$ is small compared with unity.

(6) IMPEDANCE OF CASCADED CIRCUITS

It has been shown that the carrier channel may be represented by a vector impedance, and that this impedance may be expressed in the general form $G_1 = A_1 + j_c B_1$.

It follows that the output of the carrier channel consists of the sum of two amplitude-modulated waves, the carriers being in phase and in quadrature, respectively, with the applied carrier. If the output is applied to a second network, having a vector impedance $G_2 = A_2 + j_c B_2$, and if the networks do not react on one another, the second impedance may be regarded as

acting on the two amplitude-modulated waves separately. Hence the net output will be given by

$$V_2 = V_1(A_1G_2 + j_cB_1G_2) = [(A_1A_2 - B_1B_2) + j_c(A_1B_2 + A_2B_1)]V_1$$

Thus the net transfer ratio is

$$G = (A_1A_2 - B_1B_2) + j_c(A_1B_2 + A_2B_1) \quad . \quad . \quad (20)$$

Eqn. (20) enables the impedance of composite networks to be written down immediately; for example, the impedance of the form $1/(1 + DT)^2$ [see eqn. (11)] will have a vector impedance to amplitude-modulated quantities given by

$$\frac{[(1 + j_m\omega_mT)^2 - \omega_c^2T^2] - j_c[2\omega_cT(1 + j_m\omega_mT)]}{[(1 + \omega_c^2T^2) + 2j_m\omega_mT - \omega_m^2T^2]^2}$$

(7) ROTATIONAL MODULATION

The output of a machine of the Selsyn type, when driven at constant speed, may not be represented completely by a single amplitude-modulated wave, as is sometimes assumed. A trigonometrical expression for the output is derived in Section 111 of Reference 4. This rotational modulation may be conveniently expressed as a vector of the type under discussion here.

Assume that a constant voltage $\hat{V}_1 \cos \omega_c t$ is applied to the rotor, the stator being open-circuited. If the resistance of the rotor is neglected, the flux linkage with the rotor is

$$\psi_1 = \frac{\hat{V}_1}{\omega_c} \sin \omega_c t$$

The flux linkage with the stator is, in general,

$$\psi = \frac{k\hat{V}_1}{\omega_c} \sin \omega_c t \cos \theta'$$

where θ' is the angle between the stator and the position of maximum coupling between the stator and the rotor, and k is the effective turns ratio at maximum coupling. The voltage induced in the stator will be

$$v = \frac{d\psi}{dt} = \mathrm{D}\frac{k\hat{V}_1}{\omega_c} \sin \omega_c t \cos \theta' \quad . \quad . \quad . \quad (21)$$

If the rotor is turned at a constant angular speed, then $\theta' = \omega_m t$, and v may be expressed vectorially by

$$V = \frac{\hat{V}}{j_c\omega_c}(j_m\omega_m + j_c\omega_c)\varepsilon^{(j_m\omega_m + j_c\omega_c)t} \quad . \quad . \quad . \quad (22)$$

where $\hat{V} = k\hat{V}_1$. If the rotational effect were neglected the output would be given by $\hat{V}\varepsilon^{(j_m\omega_m + j_c\omega_c)t}$. Thus the effect of the rotational voltage is to multiply the output by a term $\left(1 - j_m j_c \frac{\omega_m}{\omega_c}\right)$ which will be approximately unity when $\omega_c \gg \omega_m$. A similar result may be obtained when the rotor oscillates with small amplitude about a mean deflection.

The instantaneous value of a rotationally modulated voltage is the sum of two amplitude-modulated voltages; therefore the vector impedance developed in the paper may be used to calculate the output of any given circuit which is fed from a Selsyn or similar device.

For example, the output of a circuit having vector impedance G $(= A + j_cB)$, when a voltage of the form given in eqn. (22) is applied to it, is

$$V_3 = \hat{V}\left(1 - j_m j_c \frac{\omega_m}{\omega_c}\right)(A + j_cB)\varepsilon^{(j_m\omega_m + j_c\omega_c)t} \quad . \quad . \quad . \quad . \quad (23)$$

$$= \hat{V}\left[\left(A + j_m\frac{\omega_m}{\omega_c}B\right) + j_c\left(B - j_mA\frac{\omega_m}{\omega_c}\right)\right]\varepsilon^{(j_m\omega_m + j_c\omega_c)t} \quad . \quad (24)$$

and the output of a detector will be

$$V_2 = \hat{V}\left[\left(A + j_m\frac{\omega_m}{\omega_c}B\right)\cos\phi_R + \left(B - j_mA\frac{\omega_m}{\omega_c}\right)\sin\phi_R\right]\varepsilon^{j_m\omega_m}$$

from which

$$V_2 = \hat{V}\left[(A\cos\phi_R + B\sin\phi_R) + j_m\frac{\omega_m}{\omega_c}(-A\sin\phi_R + B\cos\phi_R)\right]\varepsilon^{j_m\omega_m} \quad . \quad (25)$$

If the rotational voltage were neglected, the output would be given by

$$V_2' = \hat{V}(A\cos\phi_R + B\sin\phi_R)\varepsilon^{j_m\omega_m} \quad . \quad . \quad (26)$$

The rotational voltage effectively increases the output by adding a term

$$j_m\frac{\omega_m}{\omega_c}\hat{V}(-A\sin\phi_R + B\cos\phi_R)\varepsilon^{j_m\omega_m}$$

(8) AN OPERATIONAL METHOD FOR INVESTIGATING THE RESPONSE OF CIRCUITS TO NON-SINUSOIDAL AMPLITUDE-MODULATED QUANTITIES

It has been shown that the vector notation defined in Section 2 may be used to determine the response of linear networks to harmonic amplitude-modulated quantities. In Reference 8 a complementary operational method is developed for the purpose of determining the response of circuits to signals which consist of a carrier whose amplitude is modulated in accordance with any general function of time. A summary of this operational method is given in Sections 8.1 to 8.3.

(8.1) Bifid Operators

The operators $\mathrm{D_m}$ and $\mathrm{D_c}$ will be defined by the equations

$$\mathrm{D_m}(mc) = \frac{dm}{dt}c \qquad \mathrm{D_c}(mc) = m\frac{dc}{dt} \quad . \quad . \quad (27)$$

where m and c are both functions of t. The operator $\mathrm{D_m}$ represents a differentiation of the function (mc) in which c is regarded as a constant, and similarly $\mathrm{D_c}$ differentiates c but regards m as a constant.

(8.2) Some Properties of the Operators

The derivative of a product is given by

$$\frac{d(mc)}{dt} \equiv \frac{dm}{dt}c + m\frac{dc}{dt}$$

or in operational form

$$\mathrm{D}(mc) \equiv (\mathrm{D_m} + \mathrm{D_c})(mc)$$

or

$$\mathrm{D} \equiv (\mathrm{D_m} + \mathrm{D_c}) \quad . \quad . \quad . \quad . \quad . \quad (28)$$

The operations denoted by $1/\mathrm{D_m}$ and $1/\mathrm{D_c}$ denote the inverse of $\mathrm{D_m}$ and $\mathrm{D_c}$ respectively, thus

$$\frac{1}{\mathrm{D_m}}(mc) = c\int mdt \qquad \frac{1}{\mathrm{D_c}}(mc) = m\int cdt \quad . \quad (29)$$

This is verified by operating on both sides of the two equations by $\mathrm{D_m}$ and $\mathrm{D_c}$ respectively.

In general, the operators obey all the rules of algebra, i.e. functions of D, D_m, and D_c may be added or multiplied together in any order, and the binomial theorem may be used to expand functions of the operators. No general proof of this is given, but it may be verified easily in particular cases. For example,

$$D^n(mc) = (D_m + D_c)^n(mc)$$

$$= \left[D_m^n + nD_m^{(n-1)}D_c + \frac{n(n-1)}{2!}D_m^{(n-2)}D_c^2 \ldots D^n\right](mc) \quad . \quad . \quad . \quad . \quad (30)$$

This corresponds to Leibnitz's theorem.

Similarly,

$$\frac{1}{D} = \frac{1}{D_m + D_c} = \frac{1}{D_c}\left[1 - \frac{D_m}{D_c} + \left(\frac{D_m}{D_c}\right)^2 - \left(\frac{D_m}{D_c}\right)^3 + \ldots\right] \quad (31)$$

This result may be used to write down the integral of expressions such as $t^3 \sin t$ immediately. Taking c as $\sin t$ and m as t^3 we have

$$\int (t^3 \sin t)\, dt = -t^3 \cos t + 3t^2 \sin t + 6t \cos t - 6 \sin t + K$$

(8.3) The Application of Bifid Operators to Amplitude-Modulated Quantities

As an example of the use of bifid operators, let it be required to find the output of a four-terminal network having a transfer ratio $1/(1 + DT) = G$, when the input consists of a voltage $v = m \cos \omega_c t$, m being any function of time. We have

$$G = \frac{1}{1 + DT} = \frac{1}{1 + (D_m + D_c)T}$$

but the carrier may be represented by the vector $\varepsilon^{j_c\omega_c t}$, so that $D_c = j_c\omega_c$. Hence the output will be given by

$$v_2 = \mathscr{R}\left(\frac{m}{1 + j_c\omega_c t + D_m T}\right)\varepsilon^{j_c\omega_c} \quad . \quad . \quad . \quad (32)$$

The expression within brackets may be evaluated for any given value of m by means of conventional methods. For example, if $m = t$ the Laplace transform of m will be $(1/s)^2$, and, assuming zero initial conditions, the quantity within brackets is given by the inverse transform

$$\mathscr{L}^{-1}\frac{1}{[(1 + j_c\omega_c T) + sT]s^2}$$

$$= \frac{1}{(1 + j_c\omega_c T)^2}\mathscr{L}^{-1}\left[\frac{T^2}{(1 + j_c\omega_c T) + sT} + \frac{1 + j_c\omega_c T}{s^2} - \frac{T}{s}\right]$$

$$= \frac{1}{(1 + j_c\omega_c T)^2}[T\varepsilon^{-(1+j_c\omega_c T)t/T} + t(1 + j_c\omega_c T) - T]$$

and the output of the circuit will be

$$v_2 = \mathscr{R}\left[\frac{T\varepsilon^{-t/T}}{(1 + j_c\omega_c T)^2} + \frac{t\varepsilon^{j_c\omega_c t}}{(1 + j_c\omega_c T)} - \frac{T\varepsilon^{j_c\omega_c t}}{(1 + j_c\omega_c T)^2}\right]$$

$$= T\frac{(1 - \omega_c^2T^2)\varepsilon^{-t/T}}{(1 + \omega_c^2T^2)^2} + \frac{t\cos(\omega_c t - \arctan \omega_c t)}{\sqrt{(1 + \omega_c^2T^2)}}$$

$$- \frac{T\cos(\omega_c T - 2\arctan \omega_c T)}{(1 + \omega_c^2T^2)} \quad .$$

This particular problem may be conveniently solved by an alternative method, because G may be expanded by means of the binomial theorem thus

$$\frac{1}{(1 + D_cT) + D_mT} = \frac{1}{(1 + D_cT)}\left\{1 - \frac{D_mT}{(1 + D_cT)} + \left[\frac{D_mT}{(1 + D_cT)}\right]^2 - \ldots\right\}$$

from which

$$\frac{t\cos\omega_c t}{(1 + D_cT + D_mT)} = \frac{t\cos(\omega_c t - \arctan\omega_c T)}{\sqrt{(1 + \omega_c^2T^2)}} - \frac{T\cos(\omega_c t - 2\arctan\omega_c T)}{(1 + \omega_c^2T^2)} + K\varepsilon^{-t/T}$$

where K is the constant of integration. In order that the output may be zero when t is zero, the constant must be

$$K = T\frac{(1 - \omega_c^2T^2)}{(1 + \omega_c^2T^2)^2}$$

(9) ACKNOWLEDGMENTS

The work described was carried out in the Engineering Department of the University College of North Wales, Bangor, while the author was in receipt of a grant from the Department of Scientific and Industrial Research.

(10) REFERENCES

(1) MacColl, L. A.: "Fundamental Theory of Servomechanisms" (D. Van Nostrand, 1945).

(2) Bubb, F. W.: "A Network Theorem and its Applications," *Proceedings of the Institute of Radio Engineers*, 1951, **39**, p. 685.

(3) Blanton, H. E.: "Carrier Compensation for Servomechanisms," *Journal of the Franklin Institute*, 1950, **250**, p. 391.

(4) Morris, D.: "A Theoretical and Experimental Method of Modulation Analysis for the Design of A.C. Servo Systems," from "Automatic and Manual Control" (Butterworth, London, 1952), p. 521.

(5) Bojornson, G. A.: "Graphical Synthesis of Networks for A.C. Servos," *Electrical Engineering*, 1951, **70**, p. 571.

(6) Riebman, L.: "Distortion in Linear Passive Networks," *Proceedings of the Institute of Radio Engineers*, 1951, **39**, p. 692.

(7) Hollingworth, J.: "The Symbolic Method," *Bulletin of Electrical Engineering Education*, 1953, **10**, p. 22.

(8) Candy, C. J. N.: "Bifid Operators," *Mathematical Gazette*, 1954, **38**, p. 270.

(11) APPENDICES

(11.1) Complex Operators

The technique described in the paper makes use of a hypercomplex system, namely the set of all the numbers of the form

$$Y = (a1 + bj_m + cj_c + dj_p) \quad . \quad . \quad . \quad . \quad (33)$$

where a, b, c and d are any real numbers and $(\pm 1, \pm j_m, \pm j_c, \pm j_p)$ is a group of eight operators, satisfying the following equations:

(i) $(1\alpha = \alpha 1 = \alpha)$; $[(-1)\alpha = \alpha(-1) = (-\alpha)]$; for each of the eight operators, α.

(ii) $(j_m^2 = -1 = j_c^2)$; $(j_p^2 = +1)$.

(iii) $(j_m j_c = j_p = j_c j_m)$; $(j_c j_p = -j_m = j_p j_c)$; $(j_p j_m = -j_c = j_m j_p)$.

(iv) A hypercomplex number Y can vanish only if each of its four components vanishes. Thus, for a, b, c, d, real, if

$$(a1 + bj_m + cj_c + dj_p) = 0$$

then

$$a = b = c = d = 0 \quad . \quad . \quad . \quad . \quad . \quad (34)$$

Any complex number Y may be expressed in the form given in eqn. (33) by rationalizing, and this expression is unique, because if Y could also be expressed as $(a_1 1 + b_1 j_m + c_1 j_c + d_1 j_p)$ then by eqn. (34) $a = a_1$, $b = b_1$, $c = c_1$, and $d = d_1$. It follows that $\mathscr{R}(Y)$ is uniquely defined by $\mathscr{R}(Y) = a$.

The method of analysis described in the paper conforms to the following outline:

(*a*) The applied voltage (or current) $x(t)$ is represented by the real part of a complex number X, which is a function of time.

(*b*) The dependent variable $y(t)$ will be given by

$$y(t) = f(\mathrm{D})x(t) \equiv f(\mathrm{D})(\mathscr{R}X) \quad . \quad . \quad . \quad (35)$$

where $f(\mathrm{D})$ is a real function of the differential operator D.

(*c*) It has then been assumed that

$$y(t) = \mathscr{R}[f(\mathrm{D})X] \quad . \quad . \quad . \quad . \quad . \quad (36)$$

This assumption is justified because X can be uniquely expanded in the form

$$X = x(t)1 + bj_m + cj_c + dj_p$$

from which

$$f(\mathrm{D})X = f(\mathrm{D})x(t)1 + f(\mathrm{D})bj_m + f(\mathrm{D})cj_c + f(\mathrm{D})dj_p$$

where $f(\mathrm{D})x(t)$; $f(\mathrm{D})b$; etc., are real because $f(\mathrm{D})$, $x(t)$, b, c, d are real

so that

$$\mathscr{R}[f(\mathrm{D})X] \equiv f(\mathrm{D})\, x(t) \equiv f(\mathrm{D})\mathscr{R}(X) \quad . \quad . \quad (37)$$

(11.2) Some Limitations of the Technique

The properties (i), (ii), (iii), (iv) given in Section 11.1 ensure that the hypercomplex numbers behave like ordinary complex numbers as regards addition, subtraction and multiplication. However, the number system differs from familiar number systems in one important respect, namely that division is not always possible. This is due to the existence of exceptional numbers, called "zero-divisors." For we may have $AB = 0$ although neither $A = 0$ nor $B = 0$. For example, $(1 + j_p)(1 - j_p) = 0$, but neither factor is zero.

This might have been a serious objection, since care would need to be taken in cancelling operations; e.g. if A is a zero-divisor ($\neq 0$), it does not follow from $AX = AY$ that $X = Y$. However, it is a fortunate circumstance that for the proposed application (namely to signals and carriers where $0 \leqslant \omega_m < \omega_c$) this situation can only occur when dealing with an h.f. transfer operator having a perfect (loss-free) anti-resonance, represented by $G = G_1(\mathrm{D})/(\mathrm{D}^2 + \omega_0^2)$, at a frequency which coincides with one of the sidebands $\omega_0 = (\omega_c \pm \omega_m)$, and under these conditions the analysis would fail in any case. Even if such a situation occurs disguised, it will reveal itself, during the rationalizing procedure, by a zero in the denominator.

It should be noted that care must be taken when it is necessary to find the root of a hypercomplex number. The following are examples:

If $Y^2 = a1$ then $Y = \pm(\sqrt{a})1$ or $\pm(\sqrt{a})j_p$

If $Y^2 = -a1$ then $Y = \pm(\sqrt{a})j_m$ or $\pm(\sqrt{a})j_c$

If $Y^2 = bj_m$ then $Y = \pm\left(\sqrt{\frac{b}{2}}\right)(1 + j_m)$ or $\pm\left(\sqrt{\frac{b}{2}}\right)(j_p - j_c)$

If $Y^2 = dj_p$ then $Y = \pm\frac{\sqrt{d}}{2}(1 + j_m)(1 + j_c)$

or $\pm\frac{\sqrt{d}}{2}(1 - j_m)(1 - j_c)$

DISCUSSION BEFORE A JOINT MEETING OF THE MEASUREMENT AND CONTROL SECTION AND THE RADIO AND TELECOMMUNICATION SECTION, 29TH NOVEMBER, 1955

Dr. David Morris: The author has followed a rather unusual procedure at this meeting, because his introductory remarks appear to have been devoted as much to his next two papers as to the two which we already have before us. I hope that we shall soon see a more complete description of the thermistor oscillators which he has just illustrated, because the very interesting arrangements which he has used to obtain the independence of the controls afford a useful example of the thermistor-circuit arrangements that can be used in practical construction. I hope also that we shall see published a more complete correlation between the hypercomplex notation that he has used in his present paper on modulation analysis, and the matrix notation that he has mentioned in the presentation.

With reference to the paper on modulation analysis, the fact that the operators appropriate to each section of the carrier channel can be combined by the cascading rules constitutes a very real advance.

In Section 13 of the paper on thermistors, thermistor circuits are compared to conventional *RCL* circuits, but the comparison is, of course, not one of exact analogy. Nevertheless, there are circuits, such as the parallel-T thermistor circuit shown by the author during the presentation of the paper, that are of a resonant nature, and their performance can be described in terms of a quality factor Q.* Now conventional passive *RC* circuits (or passive *RL* circuits), to which I think these thermistor circuits are more analogous than to *RCL* circuits, are, I believe limited to Q-factors of less than 0·5. I think it is a point of interest that we should know whether Q-factors greater than 0·5 can be realized with these thermistor circuits, in view of the fact that a negative-resistance element is present.

It is mentioned in Section 8 that the time-constant of the thermistor can be modified by changing the value of S in the quotient S/K, where S is the thermal storage constant and K is the dissipation constant. There might be an interesting field for development in reducing the value of the time-constant by using artificial ventilation to increase the value of the dissipation constant K. This might bring the thermistor of the smaller sizes mentioned by the author within the range of time-constants appropriate to mechanical control systems. As the change in the dissipation constant would be dependent upon ventilation and not upon the actual temperature, circuits might be devised with resonant frequency giving a measure of ventilation as distinct from temperature.

In Section 6, the author has described one of the techniques which he has used for measuring the parameters of these thermistors. I should like to mention one of the other methods that he has used, since it may be of wide application for these low-frequency circuits. It is possible to record the performance of thermistors by means of pen-recorders, but the commercially available pen-recorders have a paper speed that is too high for this work. However, if a recorder with a synchronous motor

* 'Q as a Mathematical Parameter', *Electronic Engineering*, July, 1954.

drive is employed, it can be powered from a 3-phase alternator driven at an adjustable low speed. The alternator can conveniently be of a small permanent-magnet-excitation type. It can then be geared to the same shaft as the modulator. In this way the paper speed is matched to the frequency, so that we have the advantages, not only that the paper speed is reduced to a convenient value, but also that the wavelength of the trace as recorded on the paper is of constant length for all test frequencies.

Dr. J. H. Westcott: There are one or two things which struck me as unfamiliar on seeing this work for the first time. For instance, the equivalent circuit with a negative resistance is quite correct, but I think it would not have surprised me so much if it had appeared as a balanced, or even an unbalanced, bridge. It might be so represented up to the point at which it becomes unstable, and after that it is of very limited application anyway.

In the Appendices there is a note on the temperature dependence of thermistor characteristics, and I have been thinking how this manifest weakness could be overcome. My first scheme—which I believe has been done already—was to embody a little heater so that the base heating was fairly high to start with and thus more or less independent of ambient changes. This seems feasible provided that it is not carried so far that it radically alters the characteristics.

But, following a line a little nearer home, suppose we agree to take the characteristics as they come, could we not do quite well by using feedback techniques, because we are quite used to dealing in this way with elements whose gains are notoriously variable? These variations in this case are rather considerable, I admit. In this respect the polar diagrams interest me very much. For changes of the polarizing current, which must be very nearly related to temperature, the shape of the frequency response stays the same and only the gain changes; this can be stabilized by using standard feedback techniques in the usual manner, exchanging gain for stability.

I thought the practical forms in the Section on the phase-shift circuits were not really very practical. They would work, but it seems that some tidying up needs to be done. I am a little distressed to find polarizing batteries used freely, isolated and very awkwardly. Since it is claimed that the thermistor behaves the same way to a.c. polarization, one could more conveniently employ polarizing windings from a transformer, not unlike the customary heater windings.

There is some degree of similarity between thermistors and transistors. But here again these elements were, in the early days, notoriously variable with temperature, but ways have been found of cushioning against these variable parameters, so it seems to me that, if the author himself and some circuit genius could get together, they might produce very practical schemes with none of the objections that one can put against the present circuits. I should like to see this happen because, as circuit elements thermistors allow an extra degree of freedom—one we have certainly been wanting, particularly in the matter of a.c. modulated signals.

I gather from the second paper that the reason we have this rather pretty hypercomplex variable work is that it does offer possibilities of analysing a.c. modulated signals which could be manipulated by thermistors. I find the author's algebra very intriguing, and I wonder whether it could be extended to the hypercomplex variable—not only the hypercomplex function but the hypercomplex variable also. This would seem to be a fascinating possibility for the mathematicians, if not for us, and I see no reason why one should not go so far as to have the equivalent of the root loci methods, but applied to carrier-modulated signals.

Mr. P. P. Eckersley: The susceptibility of the characteristics of a thermistor to ambient temperature changes may lead to instabilities of a formidable nature.

For example, if a manual gain regulator is varied by the simple means of controlling a thermistor heater current, changes in the attenuation due to variations of temperature from, say, 20° to 70° C could amount to the order 12 dB.

It is common practice to connect a thermistor in the negative feedback path of an *RC* oscillator; unless the effect of ambient temperature variation is guarded against, variations in it can cause relatively large fluctuations of frequency. In a particular equipment, and without means to combat the error, an oscillator could be seen to vary by 500 in 200 000 c/s when a laboratory window was left open.

I had thought to catch the author in error in neglecting the effects of ambient temperature upon thermistors, but found on a closer examination of the paper the statement 'the effect of temperature changes can be seen from the static characteristics in Fig. 21. These are plotted for various ambient temperatures. The values of R and R_0 may be determined from these curves for any given polarizing current. It is seen that, in general, a change of temperature will result in a change in R and R_0, but the time-constant will remain unaltered.'

May I ask by how much it is 'unaltered'?

The device ought to have application for testing servo-mechanisms in relation to hunting where a very-low-frequency source is usually employed as a stimulus. It is possible to use *RC* oscillators for this purpose, but their design—notably with respect to decoupling—may present problems which, if not insoluble, at least demand rather clumsy techniques.

Mr. F. C. Widdis: The paper opens up many attractive possibilities, but it is essential that if the thermistor is to be used for the purposes suggested, its characteristics should be reproducible and free from ageing effects. Very little information is available on this point. I have carried out some experiments which indicate that, for operating temperatures up to 120° C, short-term ageing effects are undetectable and the characteristics are quite reproducible.

There seems to be some confusion in the paper over time-constants, the thermal time-constant being used in eqn. (19), where I believe the electrical time-constant should be used. This requires clarification. I have measured the thermal time-constant of a B 5412/60 thermistor and obtained a value of 9 sec. The value of 4·4 sec quoted in the paper is the electrical time-constant.

I should like to draw the author's attention to a recent paper by Burgess,* who gives a very elegant proof that the impedance locus of the thermistor is a circle.

The author's treatment appears to be based upon the a.c. signal being small compared with the polarizing current. I infer that with large signals the characteristics will change, and this may be extremely objectionable if they are used in phase-shift networks for the stabilization of systems in which large disturbances can occur.

The change in thermistor resistance with ambient temperature may be a nuisance. I believe this could be readily corrected by using an indirectly-heated thermistor as the sensitive element. The bead temperature can be artificially raised by energizing the heater at a low level, and then controlled by incorporating a thermistor in the heater circuit.

Mr. H. V. Harley: I feel that to call the thermistor an inductance, at some frequencies, is not quite correct, although we do know the voltage and current may be in quadrature.

I should like to ask the author: first, whether thermistor characteristics are sufficiently standard to enable circuits to be

* BURGESS, R. E.: 'The A.C. Admittance of Temperature-Dependent Circuit Elements', *Proceedings of the Physical Society*, 1955, 68 B, p. 766.

esigned around given types; secondly, whether he can tell us nything about superior or inferior characteristics of indirectly-eated thermistors; and thirdly, whether, because of the negative esistance which has been mentioned, there is any particular tability problem associated with the design of cascade filter etworks for very low frequencies.

Lastly, in some circumstances it is possible for valve oscillators o go into a state of self-modulation. Could one design self-nodulating oscillators working at, say, some low frequency, the nain oscillation being of conventional magslip carrier frequency, vhilst the modulation would be due to thermistor elements in he circuit? One might perhaps have a thermistor 3-phase nodulating system leading to simulation of the output voltages of a rotating magslip for testing magslip circuits.

Mr. F. J. Hyde: I should like to make a few comments on the irst paper, in connection with the derivation of the analytical form of the impedance.

In the expansion of eqn. (8), first-order terms involving ΔR nave been omitted, without apparent justification. Further, it must be recognized that ΔR is an alternating component of resistance, cophasal with the alternating component of the emperature [through eqn. (1)], but not cophasal with the alternating component of the current, except at frequencies low compared with the reciprocal of the thermal time-constant.

The analysis of Section 4.1 leads to eqn. (18), which may be converted into a 'practical' form of the impedance, involving R_0, R_∞ and τ, by differentiation of the steady-state versions of eqns. (1) and (9), bearing in mind that $R_0 = -\partial v/\partial i$ and $R_\infty = v/i$. We then have

$$Z = R_\infty - \left(\frac{R_\infty + R_0}{1 + j\omega\tau}\right)\left(\frac{2R_\infty}{R_\infty - R_0}\right)$$

vhich is seen to differ from the empirical equation (19).

A general analysis of the small-signal impedance of temperature-ensitive circuit elements has recently been carried out by Burgess.* Adapting this for the special case of a thermistor and assuming that heat is lost according to Newton's law, we may derive

$$Z = R_\infty \cdot \frac{R_\infty + R_0}{1 + j\omega\tau\left(\frac{R_\infty - R_0}{2R_\infty}\right)}$$

In a recent paper† these various expressions have been compared with experimental values for a thermistor Type A.5412. The Burgess expression has been found to give good agreement, except for conditions of high power-loading under vhich Newton's law is less likely to be valid, whereas the practical forms of eqns. (18) and (19) do not. The values derived using eqn. (19) do lie on the experimental (semi-circular) locus of the tip of the impedance vector, but are displaced along t from the individual experimental points; they can be brought nto coincidence if τ in eqn. (19) is replaced by $\tau(R_\infty - R_0)/2R_\infty$.

Mr. F. J. U. Ritson: Reference has been made to networks ncorporating thermistors which are able to produce phase-dvance effects in an a.c. data transmission system. These etworks would appear to be potentially valuable for the stabiliza-ion of a.c. servo mechanisms, because the former are relatively imple structures as compared with the carrier filters normally sed for this purpose. Furthermore, the thermistor network has n advantage in that it does not require the maintenance of an ccurate carrier frequency for its proper operation, nor does it mploy close tolerance components.

For normal instrument-size servo mechanisms, undamped natural frequencies lie between the limits 1–10 c/s. If these servo mechanisms were to be stabilized by 'phase advance,' roughly speaking, effective time-constants of from 0·15 to 0·015 sec would be required in an ideal phase-advance network to provide adequate damping.

The thermistors used by the author in his phase-advance network appear to give time-constants too large by about one order of magnitude to be useful in this application. It would be interesting to know whether there are any fundamental difficulties in producing thermistors with very much shorter time-constants than the ones already available, since their manufacture would appear to be worth while.

Prof. A. Tustin: During the 1939–45 War I carried out some preliminary experiments on the use of thermistors as phase advances for the stabilization of servo systems. The point of view then adopted was a little wider than that taken in the present paper. We saw the problem in terms of arranging a simple feedback, in the form of a back e.m.f., produced with a time lag. Nothing seemed to be easier than to have a temperature-dependent resistance carrying a standing current and to heat it up sufficiently slowly in dependence on an output current. We thought in terms of indirect heating. The additional voltage which is caused in the circuit by the change of resistance may be given various time-characteristics. For instance, the heating may be via one or more intermediate parts, each having thermal capacitance, so that one has the equivalent of a ladder network.

If indirect heating is used, one is not worried by any question of the interconnection between the circuits or by the phase if the heating current is an alternating carrier.

The development was not followed up for two reasons: one, the effect of the ambient temperature, and the other, the fact that it takes some time for these devices to get into their steady working state. The latter feature is important where the time-constants involved are large, as in applications in chemical plant. Otherwise, thermistor devices with indirect heating are particularly inviting when large time-constants are required.

After these early experiments some work was carried out at my suggestion at King's College, where some models were made using indirect heating, with the idea of application to control in the chemical industry. It would be interesting to know the results of this work.

Prof. R. E. Burgess (*Canada: communicated*): In the analysis of the small-signal a.c. impedance of a thermistor when polarized with d.c. bias, the neglect of ΔR in eqn. (8) puts the analysis in error; it may be corrected in eqns. (12) onwards by multiplying K and dividing τ by the factor

$$1 - \frac{\alpha\bar{R}i^2}{K} = 1 + \frac{b\bar{R}i^2}{K\bar{\theta}^2} = \frac{2R_\infty}{R_\infty - R_0}$$

The correction is therefore important when the d.c. bias is sufficient to bring the operating point to the non-ohmic part of the static characteristic. A general analysis leading to the a.c. impedance and equivalent circuits which is also applicable to non-linear temperature-dependent circuit elements has recently been given.* The comparison† of the rather similar static current/voltage characteristics of a thermistor and a point-contact diode reveals important basic differences which demonstrate that the thermal analysis is inadequate for diodes; thus, unfortunately, the use of the latter as short-time-constant thermistors is not possible.

* Burgess, R. E.: 'The A.C. Admittance of Temperature-Dependent Circuit Elements', *Proceedings of the Physical Society* B, 1955, **68**, p. 766.
† Hyde, F. J.: 'Impedance of a Thermistor at Low Frequencies,' *Journal of Electronics*, 1955, **1**, p. 303.

* Burgess, R. E.: 'The A.C. Admittance of Temperature-Dependent Circuit Elements', *Proceedings of the Physical Society* B, 1955, **68**, p. 760.
† Burgess, R. E.: 'The Turnover Phenomenon in Thermistors and in Point-Contact Germanium Rectifiers', *ibid.*, p. 908.

THE AUTHOR'S REPLY TO THE ABOVE DISCUSSION

Dr. C. J. N. Candy (*in reply*): I agree with Dr. Morris that the time-constant may be reduced by increasing the dissipation. However, this would result in a fall in the sensitivity of the thermistor, and R_0 would approach $-R_\infty$ when K became very large. This may be an advantage in some applications, since a positive low-frequency impedance would remove the danger of circuit instability.

The Q-factor of the circuits described must be less than 0·5 as a condition of stability. However, thermistor circuits incorporating a capacitor may have very large Q-factors.

Dr. Westcott and Mr. Widdis have suggested methods of compensating for the variations in ambient temperature by means of a small heater; an alternative scheme would be to control the polarizing current.

I have used a thermistor phase-advance circuit in a position-control system; it remained stable over a wide temperature range, despite the changes in the thermistor impedance. The difficulty arising from the non-linear resistance of the thermistor to large signals, mentioned by Mr. Widdis, was overcome by ensuring that under maximum error conditions the net current in the thermistor was greater than that which produces the voltage maximum in the characteristics.

The polarizing batteries to which Dr. Westcott objects may all be replaced by resistance networks connected to positive and negative high-voltage supplies. They may be replaced by transformers only when modulated signals are employed.

Both Mr. Widdis and Mr. Harley question the consistency of the characteristics: I have found them to be as good as those of transistors. Mr. Harley also mentioned the problem of stability, which must be an important aspect of any design method. A circuit incorporating two thermistors is stable when the resistance presented to one thermistor is greater than R_0, provided that the circuit is stable when the incremental impedance of this thermistor is considered to be zero (see Reference 10).

An oscillator incorporating a.c. polarized thermistors has been designed. It would, however, be difficult to generate a variable-frequency 3-phase modulated oscillation with a fixed phase difference between the three voltages, as Mr. Harley requires.

I am indebted to Mr. Hyde for first pointing out the discrepancy between my result and that obtained by Prof. Burgess. This error may be corrected by modifying K as Prof. Burgess suggests. The fact that ΔR is not in phase with i is in accordance with my result.

The error does not invalidate the design technique, provided that the time-constant is given by

$$\tau = \frac{S}{K}\left(\frac{R_\infty - R_0}{2R_\infty}\right)$$

where τ is the electrical time-constant as measured in Section 5, not the thermal time-constant assumed by Mr. Hyde. The above expression shows the variation in τ as the polarizing current and the temperature change. It appears that τ could be brought into the range defined by Mr. Ritson by making $R_0 \simeq R_\infty$.

Prof. Burgess has found a difference between the point-contact diode and the thermistor; this may be due to the fact that the diode is connected to a comparatively large metal base, in order to obtain maximum dissipation, and it is probable that K is a function of frequency in these conditions.

I would like to pay tribute to Prof. Tustin's early work on indirectly heated resistors. The indirectly heated thermistor is equivalent to an impedance Z_T in series with a voltage $I_h R_x/(1 + D\tau)$. These thermistors do not fit well in valve circuits, for they require relatively large currents and will not withstand large voltages; however, they combine ideally with transistors.

The hypercomplex variable mentioned by Dr. Westcott may be used to solve differential equations having a time-varying parameter.

21.317.74: 621.315.61:538.566

Paper No. 2008 R
May 1956

A CENTIMETRE-WAVE PARALLEL-PLATE SPECTROMETER

By P. H. SOLLOM, O.S.B., B.Sc., Ph.D., and J. BROWN, M.A., Ph.D., Associate Member.

(*The paper was first received 24th August, and in revised form 29th November*, 1955.)

SUMMARY

A spectrometer has been designed for the measurement at a wavelength of 1·25 cm of the refractive index of samples formed into prisms. The radiation is totally enclosed by two parallel plates approximately 4 ft in diameter, $\frac{3}{16}$ in apart, and enters the instrument by a sectoral horn free to move over a 90° sector of the circumference. The receiving aerial can be either a similar horn or an open-ended waveguide, and may be positioned anywhere on the circumference. The principles on which the instrument is designed are discussed, and details are given of the mechanical construction and of the tests made upon its accuracy. The electrical performance has been examined in a variety of ways, and it has been established that beam directions can be located to within 0·25°. Variations in the received signal power occur and are shown to arise from a variety of effects, such as plate spacing, aerial penetration into the plate region and interference by unwanted signals such as the transmitter side-lobes. None of these causes a significant change of the direction of the main beam. Measurements on polystyrene prisms show a 3% error in refractive index attributed to imperfect contact between the dielectric of the prism and the spectrometer plates. This error can be reduced to less than 0·5% by coating the prism surfaces with tin foil.

(1) INTRODUCTION

At wavelengths less than 3 cm, optical methods may be used to determine the electrical properties of materials in place of the waveguide methods commonly used at the longer wavelengths in the microwave region. There are two general types of optical method for finding the permittivity of a dielectric, the first relying on a direct measurement of the electrical path-length through a parallel-sided sample, and the second on measuring the change in direction of a plane wave which is made to pass through a prism. The optical instruments used for such measurements are the interferometer for the first method and the spectrometer for the second. Both instruments may be designed to operate at centrimetric wavelengths and a number of examples have been described by Culshaw.[1,2,3]

In a spectrometer, the direction of propagation of the radio waves may be easily changed, allowing a possible dependence of the electrical properties of materials on direction of propagation to be investigated. The instrument described in the paper has been designed with the intention of carrying out such an investigation on non-isotropic artificial dielectrics.[4] At the same time, the possibility of using it for studies of diffraction gratings and cylindrical radiators has been borne in mind and has led to the introduction of certain features to make the performance of the instrument as versatile as possible.

The essential feature of any spectrometer is provision for a beam of radiation to pass through a prism of the material under test. The direction of the beam leaving the prism is measured by a suitable detector, and the refractive index of the material can be calculated from the change in the direction of the beam and the angle of the prism. In optical instruments, the dimensions of the radiating and detecting apertures and of the prism are all very much larger than the wavelength of light, so that the light travels in parallel-sided beams and its behaviour may be predicted from geometrical optics. Even at the shortest radio wavelengths now available the corresponding dimensions are effectively restricted under laboratory conditions to a range of from ten to one hundred wavelengths and the radiation no longer behaves in such a simple manner. As is well known from studies of microwave aerials, the beams diverge instead of being parallel-sided, the angular spread being inversely proportional to the number of wavelengths in the radiating aperture. Despite this complication, useful results may be obtained from spectrometers of small size compared to the operating wavelength. The most direct adaptation of optical design requires radiating and receiving aerials with apertures at least ten wavelengths square and a prism whose surfaces are of comparable size. The radiated beam travels from the transmitter aerial to the prism and from the prism to the receiving aerial under free-space conditions. Instruments of this type have been described by Cochrane,[5] Culshaw[3] and Brady, Pearson and Peoples.[6]

Artificial dielectrics consist of regular arrays of conductors supported by a solid or expanded plastic, and a considerable amount of time is often required for the construction of samples. A desirable feature of any measurement technique is therefore that it should require only a small sample. The type of spectrometer described above is unsuitable from this point of view and a modified form, in which the radiation is confined within the region bounded by two parallel conducting plates, has therefore been developed. The wave within this region is identical to a plane wave in air provided that the direction of the electric field is perpendicular to the plates. The height of the prism required in a spectrometer of this type is the same as the spacing between the plates, which must not exceed one wavelength if trouble from the undesired modes of propagation is to be avoided. The prism height is therefore at most only one-tenth of that required in the simple spectrometer, leading to a considerable reduction in the labour of preparing sample prisms. Transmission lines based on the parallel-plate principle have already been used to measure the reflections at artificial dielectric surfaces[7] and to examine the fields diffracted by cylindrical obstacles, while an elementary form of parallel-plate spectrometer has been described by Ruze and Young.[8]

(2) GENERAL DESCRIPTION OF THE INSTRUMENT

Most of the problems likely to be investigated with the help of the spectrometer are such that the absolute wavelength used is unimportant, the critical quantity being the ratio of the wavelength to some dimension of the test sample, such as the lattice spacing of the conductors in an artificial dielectric. The use of a very short wavelength minimizes the spreading of the radiated beam, while longer wavelengths lead to less severe tolerances on sample dimensions. The first consideration is the more pressing, and a choice of 1·25 cm was made for the operating wavelength, this being the shortest for which a supply of waveguide components and oscillators was readily available when the construc-

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Dr. Sollom, who was formerly in the Electrical Engineering Department, Imperial College, University of London, is now at Douai Abbey.

Dr. Brown is in the Electrical Engineering Department, Imperial College, University of London.

tion of the instrument was commenced. The supply of oscillators for this wavelength in the future appeared somewhat uncertain, and the instrument was therefore so designed as to make possible a change of operating wavelength to 8 mm with the minimum of expense and labour.

The principal components of the spectrometer are the parallel plates and the transmitting and receiving aerials. The plate spacing is a little larger than the smaller dimension of the transmitter waveguide cross-section, making sectoral horns flared in the magnetic plane a suitable choice for the aerials. A crystal diode connected directly to the receiving aerial forms an adequate detector, and since the rectified output may be fed through a coaxial cable no difficulty arises in moving the receiving aerial. At least one additional movement is required and it may be either the prism under test or the transmitting aerial. The possibility of including a turntable at the centre of the instrument was considered but rejected because of the inevitable discontinuity at the turntable circumference. In principle, this discontinuity could be made electrically harmless by a choke joint of the type used in waveguide couplers, but the construction of such a joint would be extremely difficult. Further, the restricted range in operating wavelengths would necessitate a new joint if the change to a wavelength of 8 mm should be made. Provision for moving the transmitter aerial is thus included and is one of the features which allows the use of the spectrometer for measurements other than those on prisms. The attenuation in coaxial cable at a wavelength of 1·25 cm is too large to permit its use as a connection between the oscillator and the transmitting aerial. The other possibilities are a flexible waveguide and a solid waveguide run including a rotating joint; the solid waveguide was selected as the more reliable in operation.

These general considerations suffice to define the essential characteristics of the instrument. The general layout takes the form shown in Fig. 1. This is a photograph of the completed spectrometer with the upper plate raised, and it shows a test prism lying on the lower plate. When measurements are being made, the upper plate rests on absorbing spacers positioned around the circumference. Both aerials are mounted on carriages which are supported by rollers running on the lower plate and by arms connected to bearings on the main supporting pillar. The rotating joint for the transmitter is mounted inside this pillar, access being by windows which permit movement through an arc of 90°.

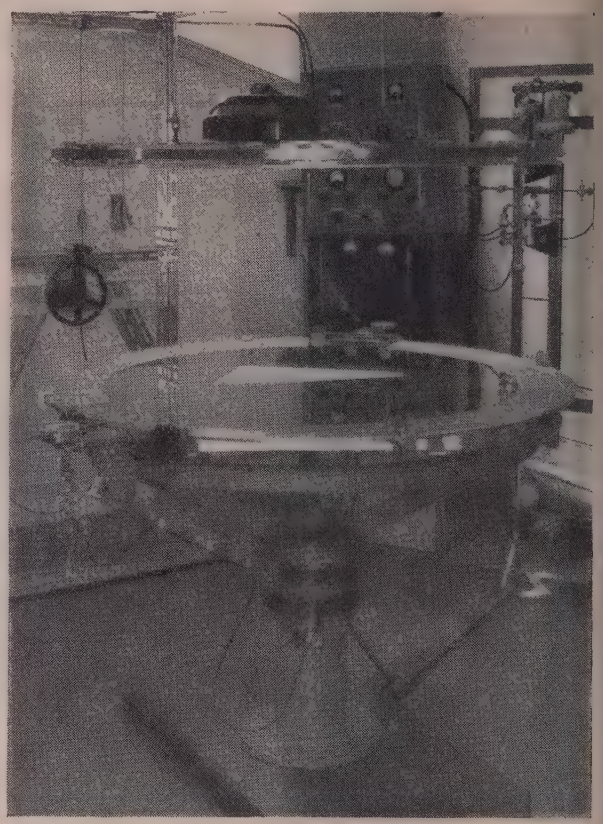


Fig. 1.—General view of completed spectrometer.

The top plate is shown in the raised position for the insertion of test specimens.

(3) MECHANICAL DESIGN

The required mode of propagation within the parallel plates is a transverse electromagnetic wave which behaves in the same way as a plane wave in free space only if the plates are accurately plane and parallel. In practice, gradual departures from this ideal have very little effect on performance, but sudden discontinuities may cause partial reflection of an incident wave. For this reason it was decided to make each plate from a single sheet of metal, thus avoiding the step which must occur at the junction between two pieces. The largest stock size of metal sheet, namely 8 × 4 ft, limits the possible diameter of the plates to 4 ft. The diameter of the top plate was selected to make the circumference exactly 360 cm, thus allowing an angular scale to be provided by four steel rules each 90 cm long, and is therefore approximately 3 ft 9 in. The rules are marked in millimetres, permitting angular measurements to a tenth of a degree. The diameter of the lower plate is 3 ft 11 in, leaving a lip one inch in width along which the aerial carriage rollers may run.

The plates were made from aluminium sheet $\frac{1}{4}$ in thick, shaped approximately by a band saw and then accurately filed to the required diameters. A stiffening framework is essential to keep each plate flat and is provided in the form of an interlocking triangular structure made from $1\frac{1}{2} \times 1\frac{1}{2} \times \frac{1}{4}$ in T-section aluminium. The aluminium members were cut to the correct lengths and welded at the intersections. During the welding process the framework was distorted by the heat, but was subsequently machined flat while mounted on a lathe face-plate. Welding each plate to its supporting structure would have caused too great a distortion, and to avoid this the plates were secured to the frames by self-tapping screws. Hardened self-tapping screws were driven through the framework and into blind holes drilled in the plates to a maximum depth of $\frac{3}{16}$ in. This has had no visible effect on the flatness of the plates.

The bottom plate is supported on six adjustable screws carried on an angle-iron ring, itself fixed to the central pillar by four spokes of T-iron. The plate is levelled on three of these screws and is positioned so that its centre lies on the pillar axis. The remaining screws are then raised to touch steel cushions on the framework and to support the weight at these points.

The top plate may be lifted by the simultaneous movement of three cables passing over pulleys carried on a gantry constructed from aluminium girders. Steel locating pins are fixed at the zeros of the four angular scales, and when the top plate is lowered these pins engage in V-shaped guides clamped to the bottom plate and accurately positioned at the ends of two perpendicular diameters. These guides are intended only to position the top plate, the weight being carried by spacers positioned around the circumference, and when measurements are being made the guides can be removed from the instrument.

Each horn carriage is mounted on a radial arm attached to a bearing on the central pillar and has three rollers, two above and

ıe below the bottom plate. The weight of the carriage is carried nost entirely by the radial arm, the rollers being provided to ve a friction drive for moving the carriages round the spectro- eter circumference and also to allow adjustments of the horn ositions as described in Section 6.1.

The plate spacing is maintained at the selected value of $\frac{3}{16}$ in sectoral wedges of a leaded-asbestos material which behaves a radiation absorber. The spacers have a taper length of in and were made by turning down a 4 ft-diameter sheet, the sulting annulus being cut into sectors of convenient sizes. The edges form very good matched loads and absorb the radiation cident upon them. Equally, the wedges prevent stray radiation utside the spectrometer from entering the plates, so that no ouble is experienced from reflections by the laboratory walls s often happens with the more conventional forms of spectro- eter. The spacing of $\frac{3}{16}$ in is larger than the internal dimen- ons of the waveguide cross-section by about 0·017 in to allow ıe sectoral horns to penetrate within the plate region without aking contact with the plates.

(4) ELECTRICAL DESIGN

The components which require most consideration are the lectromagnetic horns used as the aerials for the transmitter and eceiver. The size of the aperture must be sufficiently large to ive a reasonably narrow radiated beam but must not occupy oo large a sector of the spectrometer circumference. The value elected is 5 in, or approximately ten wavelengths at the nominal perating frequency. The sharpest beam obtainable in practice rom such a horn requires that the phase front be flat at the perture; a polystyrene lens is therefore placed at the aperture o convert the cylindrical phase fronts in the horn into planes. his lens is designed according to the usual geometrical optical rinciples,[9] the inner surface being a hyperbolic cylinder and the uter surface a plane. Unwanted side-lobes in the radiation attern may arise as a result of reflections at the lens surfaces, but ıey can be reduced by matching one of the surfaces. This has een done by adding to the outer surface a quarter-wave trans- ormer consisting of slots of suitable width and depth machined rom the solid dielectric. The values of the slot dimensions, alculated from the results of a theoretical study of slotted ections, are 0·099 in for the depth, 0·125 in for the width and ·191 in for the centre-to-centre spacing.

Since the horn is flared in the magnetic plane, the amplitude istribution across its aperture should approximate to a cosine urve. The radiation pattern calculated according to hypothesis, or an aperture of ten wavelengths width has a half-power beam idth of 7° and a maximum side-lobe level 23 dB below that of ıe main beam.

The discontinuity between the horn and the spectrometer is ast if the horn surfaces are in perfect contact with the spectro- ıeter plates. In practice, however, the continual movement of ıe horns during measurements would soon produce wear on the rfaces in contact, leading to erratic operation, and a non- ontact arrangement was therefore decided on, as mentioned in ection 3. The form of construction selected for the horns nsures as great a clearance as possible between the horn surfaces nd the plates; the portions of the horn surfaces which overlap ıe plates are made of 0·001 in tinfoil, supported by the rigid rass bars used for the horn sides and also by the dielectric ns. A general view of one of the horns (the transmitter) ıounted on its carriage is shown in Fig. 2. The two upper iction rollers can be seen resting on the lower plate; the approxi- ıate position of the upper-plate circumference when the instru- ıent is closed is indicated by the dotted line. The shape of the ns including the slotted outside surface is also evident in the iagram. The pointer moves against the angular scale mounted

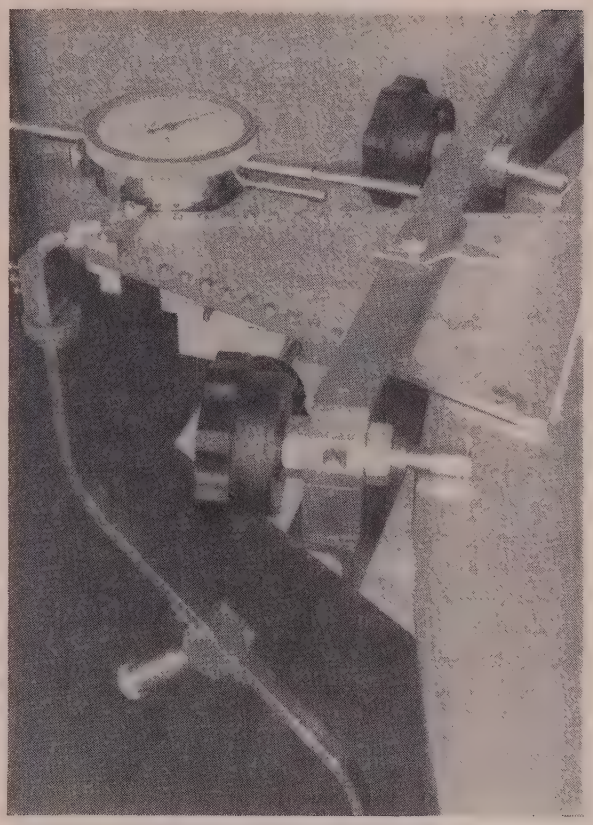


Fig. 2.—View of transmitter horn mounted on its carriage.

on the top plate. When the top plate is lowered, a cross-section through the horn axis appears as shown in Fig. 3; the friction rollers can be adjusted for positioning the horn symmetrically with respect to the plates, while the length, l, of the penetration gap can be changed by means of a rack and pinion. During measurements, any changes in the penetration gap length can be

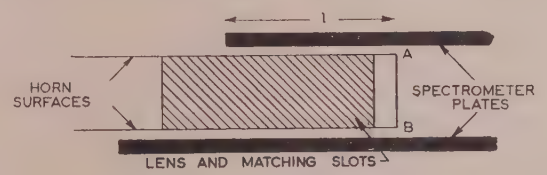


Fig. 3.—Cross-section through horn axis.

accurately obtained from the dial gauge, which is fixed to the horn and has its plunger bearing against a fixed portion of the carriage.

The complete microwave circuit of the spectrometer is shown in Fig. 4; most of the components are of standard design and need no special comment. The Pound frequency-stabilizer circuit[10] uses the modified equal-arm discriminator described by Tuller, Galloway and Zaffarano,[11] and ensures that no trouble will be experienced when tests are being made on samples whose properties are frequency-dependent. If desired, the klystron, a standard type for the 1 cm band, may be operated without a stabilizing signal, by means of the switch X. Matching screws and plungers are used liberally, since broadband matched components are not available and since it is essential to make the best

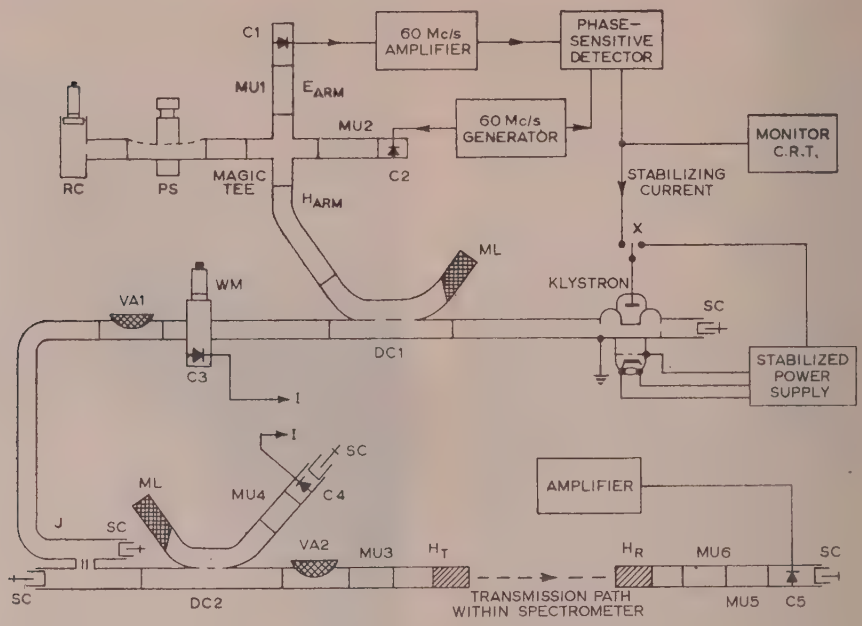


Fig. 4.—Microwave circuit for the spectrometer.

C—Crystal holder.
DC—Directional coupler.
H_R—Receiving horn.
H_T—Transmitting horn.
J—Rotating joint.
ML—Matched load.
MU—Matching unit.
PS—Phase shifter.
RC—Reference cavity.
SC—Short-circuiting plunger.
VA—Variable attenuator.
WM—Wavemeter.

possible use of the somewhat limited power output from the klystron. The crystal holder of the receiver is of a modified design which uses a window-type crystal of the 8 mm-band pattern, since this is considerably more sensitive than the usual 1 cm tubular crystal. The oscillator may be either unmodulated, the rectified current from the crystal then being measured by a galvanometer, or square-wave modulated, when the received signal is fed to a narrow-band audio amplifier.

(5) CONSTRUCTIONAL CHECKS

Before the microwave components were fitted together, a number of checks were made on the accuracy of the mechanical construction as an aid in tracing the cause of any unexpected defects in performance that might arise. The true shape and positioning of the bottom plate were checked by observing the readings on a dial gauge, which was mounted on one of the horn carriage arms and whose plunger was aligned to bear on the plate circumference. The centre of the plate was found to lie within 0·1 mm of the axis of rotation and its radius was constant to within 0·5 mm. A similar check was carried out on the top plate when it was located by the pins and V-guides, and showed the same constancy of radius; the centre positioning was correct to within 0·5 mm. A steel straight-edge, intended for the accurate positioning of specimens, was constructed and fitted with locating pins in such a way that when these pins engage in opposite V-guides the edge should lie a long a diameter. Lines drawn on a sheet of drawing paper placed on the lower plate along the edge of this steel locating jig for each of its two possible positions intersected within 0·5 mm of the centre and were perpendicular to within the limits of observation. This test not only confirmed the accuracy of the locating jig but proved that the angular scale cannot be in error by more than 0·1°, since the zeros of the four 90° sections of the scale coincide with the V-guide positions.

The flatness of the bottom plate was tested by placing a 4 ft straight-edge along a diameter and inserting feeler gauges between the plate and the straight edge. The central 6 in of the plate was found to be flat within 0·002 in and thereafter the plate fell gradually from the horizontal plane as the circumference was approached, the maximum drop being 0·012 in. Direct confirmation of the plate flatness is given by the detail visible in the reflections in Fig. 1.

(6) ELECTRICAL TESTS

(6.1) Transmitter Alignment and Matching

A comprehensive series of measurements on the behaviour of the empty spectrometer was carried out to determine how accurately the direction of a transmitted beam could be located. A first essential was to align the horns so that their electrical axes passed through the centre of the instrument, but because of their high directivity a straightforward method could not be used lest an error in the alignment of one should be annulled by an equal and opposite error in the other. This difficulty was overcome by providing a relatively non-directive aerial in the form of an open-ended waveguide, constructed with tin-foil surfaces in a way similar to the horns. This aerial was used as the receiver while the transmitter horn was being aligned. Accurate measurements of beam directions can be made only if the received signal is plotted against angular position for movements of either the transmitter or the receiver aerial. The beam direction is then taken as the visual estimate of the angle corresponding to the centre of gravity of the response. This procedure was used not only for aligning the transmitter horn but also in all subsequent measurements for which accurate beam directions were required. Once the transmitter horn was aligned, the receiver horn could be replaced and also aligned. Before this was done, however, further tests were made with the waveguide receiver and gave valuable information about the characteristics of the instrument.

At an early stage, erratic variations in the received signal were bserved and traced to radiation coming from the reflector ad of the klystron and entering the spectrometer by those parts the circumference where there were no absorbing wedges. his was cured by fitting a copper screen around the klystron, and o further evidence has been found of radiation entering the ectrometer other than by the transmitter horn.

The reflection from the transmitter-horn back along the wave- ide feed may be cancelled by the three-screw matching unit, U3, in Fig. 4. The adjustment of this unit was carried out hile the horn radiation was absorbed by a row of wedges placed arallel to the aperture; the reflected signal was observed at the ppropriate arm of the directional coupler DC2, the variable tenuator VA2 being set for zero attenuation. The match of e horn was found to remain effective for all angular settings ithin the 90° sector through which the transmitter arm may ove. The reflection from wedges placed around the spectro- eter circumference, as measured with the directional coupler C2, was found to be at least 30 dB below the level of the ansmitted wave. Removing the wedges from a 90° sector pposite the transmitter position did not lead to any appreciable eflection provided that this open sector was not obstructed in ny way. Significant reflections were observed if either a V-guide r the receiver-horn carriage was placed in this open sector, but ey could be eliminated by placing absorbing wedges directly in ont of the V-guide or in front of the carriage rollers. Very ttle power was reflected by either the waveguide or the horn ceiving aerial if the receiver crystal was adequately matched to e waveguide run.

After the transmitter horn had been aligned at one position, on the angular scale, a series of responses was taken to deter- ine whether the alignment remained constant at other angular with angular setting over a range of about 20%; the transmitted power was monitored on the directional coupler DC2 during the measurements and remained constant to within 1%.

(*d*) Although corresponding responses in the two series are not identical, there is little difference in their centres of gravity.

These measurements were made before the reflection from the receiver carriage rollers was detected; the beam displacement and distortion are partly caused by this reflection. The variations in signal level cannot be explained on this basis, and further tests were carried out to determine their cause.

(6.2) Effects of Plate Spacing and Aerial Penetration

The coupling between the aerials and the plate region must depend on the plate spacing, so that a variation of plate spacing around the circumference will cause differences in received signal level. A plot of the received power, when the two aerials were directly opposite and moved in step by intervals of 0·5°, showed fluctuations of up to 50% of the maximum value obtained. Only rough agreement was found between the power fluctuation and the measured variation in plate spacing around the circumference, showing that some other factor must also be affecting the signal levels.

A change in the aerial penetration, i.e. the distance *l* in Fig. 3, alters the effective impedances at the points A and B, and maximum received signals should be observed when these impedances are both as nearly short-circuits as possible. This state should recur at intervals of half the free-space wavelength, and a cyclic variation in signal level with this period was observed when the aerial penetration (either transmitter or receiver) was altered. A similar variation would arise if there were a standing-wave pattern in the region between the two horns, and although the observed

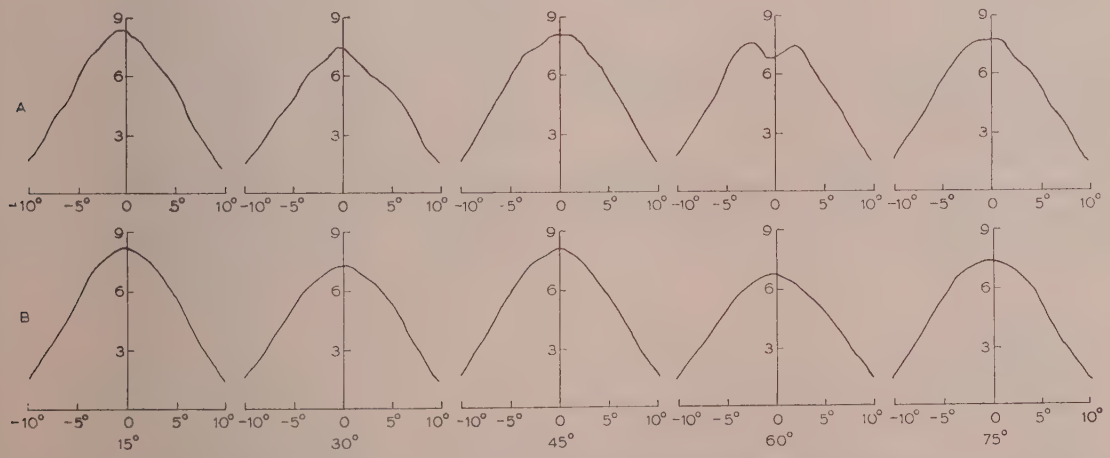


Fig. 5.—Beam responses for direct transmission through the empty spectrometer.

Each response is a plot of galvanometer deflection (vertical scale) against the angular position of the transmitter or receiver. At the angular scale zeros, the two aerials are ametrically opposite, the angle associated with each figure giving the position of the transmitter. Series A was obtained by moving the receiver only, and series B by oving the transmitter only.

ttings. These responses, which are typical of the many taken uring later work, are shown in Fig. 5. The following features ay be noted:

(*a*) The beam directions, estimated from the centres of gravity of the response curves, differ by up to 0·5° from their expected position.

(*b*) The responses are not all symmetrical and some have two maxima.

(*c*) The peak value of the galvanometer deflection fluctuates

low value of reflection from the receiver aperture (Section 6.1) suggested that this was unlikely, further confirmation was obtained by measuring the received signal as a function of both aerial penetrations at constant angular settings. The results are plotted on constant-power contours in Fig. 6; power variations arising from a standing wave would give constant power for a constant distance between the aerials, i.e. the contours would be lines making angles of 45° with both axes. There is no evidence of this. The horizontal and vertical axes drawn through the point of maximum signal cut the power contours orthogonally,

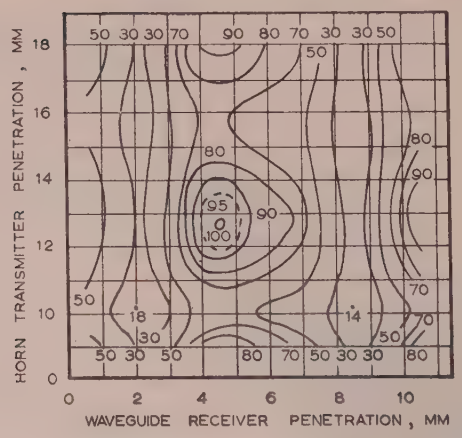


Fig. 6.—Received-signal powers as functions of aerial penetrations.

The curves are contours of constant power, expressed as a percentage of the maximum value obtained.

which show that the effects of the two aerial penetrations do not interact.

Further power contours were plotted under similar conditions to show the dependence of the received signal on the plate spacing and aerial penetration of each aerial. It was found that changes in the plate spacing and aerial penetration cannot be treated independently, a wider gap between the plates requiring a smaller penetration for maximum signal. The waveguide aerial is more sensitive to changes in penetration than the horn, possibly because of the matching layer at the horn aperture.

The interaction between plate spacing and penetration effects makes it quite impossible to explain in detail the variations in power level as the direction of transmission through the instrument is altered. However, the maximum drop in power is in good agreement with that predicted from the power contours obtained as described above for the worst measured changes in plate spacing and radius, the latter being taken to correspond to a change in aerial penetration.

The maximum rate of change of plate spacing at the circumference is 0·001 in per degree of angular movement, and this can cause a change in receiver sensitivity of 1% per degree. If the theoretical power pattern for transmission by the horn and reception by the waveguide is multiplied by this variation in sensitivity, the response maximum is displaced by 0·4°. This accounts for the beam displacements observed during direct transmission tests. When a horn is used as the receiving aerial, the displacement of the response maximum for the above rate of change of sensitivity is reduced to 0·2°. A detailed inspection of the variations in plate spacing with angular displacement, and of the dependence of the received signal upon plate spacing, confirmed the possibility of double-humped response curves. The expected dip in the maximum of response should, however, be less than that in the curve for the 60° position (Series A, Fig. 5), and, as already mentioned, the distortion in this response is attributed to reflection from the receiver aerial carriage. The tests on the empty spectrometer show that beam positions may be consistently located with errors of less than 0·1° and that departures of up to 0·5° from the theoretically predicted position may be explained by local variations in the plate configuration. A known direction error may, of course, be eliminated during future experimental work. The observed changes in the received signal power may be satisfactorily accounted for and are not likely to lead to difficulties in experiments which rely on angular measurements only.

(6.3) Measurements with a Reflecting Barrier

Since the transmitter aerial is restricted in movement to a 90° sector, the tests so far described give information on the spectrometer performance only for receiver aerial positions in the opposite 90° sector. The performance for the remaining sectors may be examined by positioning a perfectly reflecting barrier along a diameter. A suitable reflector was made by wrapping tin foil around a $\frac{1}{4}$ in-thick rectangular strip of sponge rubber; the weight of the plates compressed the rubber by about $\frac{1}{32}$ in and good contact between the foil and the plates resulted. The foil bulged outward by a small amount, but the greatest departure from the desired position of the reflecting surface did not exceed one-sixteenth of a wavelength, which was considered to be acceptable.

The reflecting barrier was placed along a diameter of the instrument and the transmitter aerial was positioned to produce a beam incident at the reflector for a range of angles of incidence. The direction of the reflected beam was determined for each angle of incidence used by plotting a response curve and was found to agree with the value expected from the optical law of reflection within the limits of accuracy determined as described in the previous Section. This confirmed that there were no additional sources of angular error in the sectors not examined in the direct transmission tests. The response curves, however, showed marked ripples with periods of 1–2°, and a rapid fluctuation in received signal level with angle of incidence was observed, as shown in Fig. 7. This curve was obtained by moving the transmitter aerial in small steps, positioning the receiver aerial at the angular position appropriate to the theoretical direction of the reflected beam, and then plotting crystal current against angle of incidence. The aerial penetrations were adjusted for maximum received signal at an angle of incidence of 12·5° and

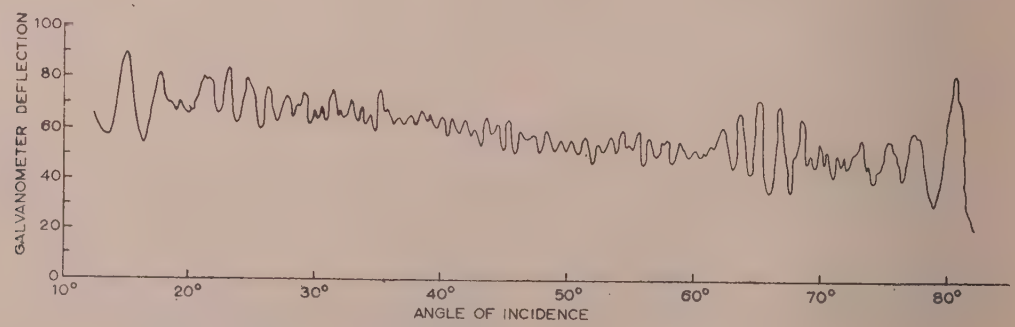


Fig. 7.—Dependence of received power on angle of incidence for reflected beam test.

eft fixed for the remaining measurements. The physical size of he aerial apertures restricted the possible range of angles of ncidence to that shown in the diagram. The mean curve drawn y averaging the rapid fluctuations shows changes in received ower of the same order as those obtained during direct trans-nission tests, and these are therefore almost certainly due to ariations in aerial penetrations and plate spacing.

The ripples in the response curve and the rapid changes in ignal level (Fig. 7) are partly explained by direct reception of a ransmitter side-lobe in addition to the main beam reflected by he barrier. Consider the arrangement of the apparatus as hown in Fig. 8. When the receiver R is moved, the path length

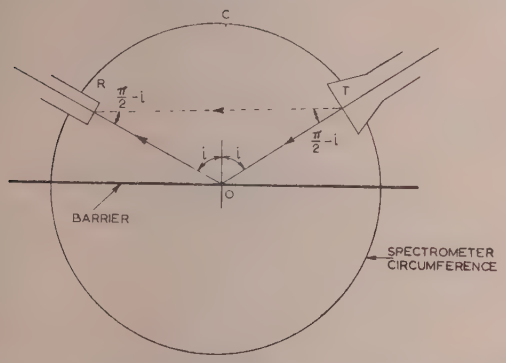


Fig. 8.—Arrangement for reflection tests, showing occurrence of interference.

ravelled by the main beam remains constant but that for the ide-lobe changes. The phase difference between the signals eceived by these two paths also changes, and successive maxima ccur for each wavelength increase in the direct path-length TR. he ripple periods in the response curves and in Fig. 7 are con-istent with this argument except at small angles of incidence. urthermore, the changes in received power in Fig. 7 give an ndication of the transmitter side-lobe levels since the waveguide eceiver is practically non-directive. A pronounced side-lobe is vident around an angle of incidence of 65°, and from the geo-netry of Fig. 8 this corresponds to a direction making an angle f 25° with the transmitter-horn axis. Direct measurement of he transmitter-horn radiation pattern showed that the second ide-lobe was at this angle and was the largest, being approxi-nately 25 dB below the level of the main beam.

Large fluctuations of received power occur at angles of inci-ence between 12·5° and 20°, although the side-lobe level for hese directions should be very low. The period of the fluctua-ons is not consistent with the side-lobe reception, and further xamination proved that in this case the interfering signal arrives y a path comprising two reflections at the barrier and one at oint C on the spectrometer circumference.

(6.4) Diffraction by a Dielectric Cylinder

If a dielectric cylinder is illuminated by a plane wave, diffrac-on occurs and the scattered field has an angular distribution milar in appearance to the polar diagram of an aerial, the umber of lobes depending on the ratio of the cylinder diameter the wavelength. With cylinders of small diameters, the ampli-de of the scattered field is small compared with that of the ain beam, so that observations of this field in the spectrometer rovide a test of its accuracy when the wanted signal is not uch larger than that from the transmitted side-lobes and stray flections. Accordingly, measurements were made with two olystyrene cylinders of diameters 0·478 and 0·796 in, corresponding to values of the parameter $2\pi r/\lambda$ of 3 and 5 respectively, r being the radius of the cylinder and λ the free-space wavelength, taken as 1·27 cm for these measurements. A cylinder was placed at the centre of the spectrometer and the received signal was plotted for all directions except those covered by the transmitter horn aperture and those lying within the range of the main transmitted beam. The scattered field from the cylinder was at least 20 dB below the level of the main transmitted beam and thus comparable in amplitude to the side-lobes. The plots of received power against the angular position of the receiving aerial therefore show ripples similar in form to those in Fig. 7, and analysis of the ripple periods gives results agreeing with those expected. Measurements on the two sides of the transmitter axis show the positions of corresponding peaks and troughs of these ripples, agreeing to within 0·5° and showing the high degree of symmetry of the instrument.

When mean curves are drawn by averaging over the ripple periods, the angular distribution of the scattered power shows the expected shape, and the positions of the minima lie within 1° of the theoretical values. But the agreement between theoretical and observed signal levels was not good, partly because of the effects already discussed and partly because the receiver horn accepts signals from a range of directions and not from one direction only. The observed received signal is therefore an average of the scattered signals over a range of angles and shows smaller differences in level between maxima and minima than are predicted theoretically.

The principal conclusions to be drawn from the experiments on the dielectric cylinders are that accurate angular measurements are possible even at signal levels of the same order as the side-lobe radiation, and that great care is necessary in interpreting the significance of changes in received powers.

(7) MEASUREMENTS ON POLYSTYRENE PRISMS

The major purpose for which the spectrometer has been designed is the determination of the refractive index of artificial dielectrics by measuring the deviation of a beam passing through a prism of the material. As a final check on the spectrometer performance, measurements were therefore made with polystyrene prisms for which the refractive index could be accurately measured in a cavity resonator. In particular, information was desired on the minimum size of prism aperture required for accurate results, and on the extent of distortion of the response curves arising from refraction by the prism.

(7.1) Theory of the Method

The simplest procedure is to position a prism with its hypotenuse along a diameter of the spectrometer and to direct the transmitter beam so that it is incident normally on to one of the other prism surfaces (see Fig. 9). This ensures that the transmitted beam passes through the spectrometer centre, eliminating complications in the interpretation of the direction of the refracted beam, and also facilitates accurate positioning of the prism with the aid of the jig described in Section 5. An application of Snell's law to the refraction at the hypotenuse gives

$$n = \frac{\sin (D + P)}{\sin P} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

from which n may be calculated if the angle of deviation, D, and the prism angle P are measured.

For a given prism, the angle P may be measured very accurately, and errors in n arise primarily from errors in estimating D from the response curves. Consideration is therefore given to the possibility of selecting the value of P to give the least error in n for the expected error in D. Differentiation of eqn. (2) with

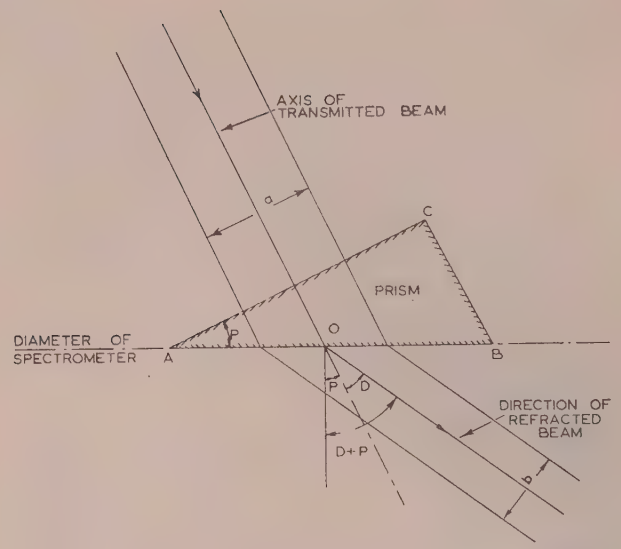


Fig. 9.—Arrangement of prism for refractive index measurement.

O—Centre of the spectrometer.
P—Refracting angle of the prism.
D—Deviation of the beam as a result of refraction.
a—Effective aperture of the incident beam.
b—Effective aperture of the refracted beam.

respect to D, assuming P constant, shows that the error δn in refractive index arising from an error δD in the deviation is

$$\delta n - \frac{\cos(D+P)}{\sin P}\delta D$$
$$= (\mathrm{cosec}^2 P - n^2)^{\frac{1}{2}}\delta D \quad . \quad . \quad . \quad . \quad . \quad (2)$$

This suggests that δn decreases as P increases. In practice, however, δD is not independent of P, first because of the magnification of the prism and secondly because of changes in the received power resulting from reflections at the prism surfaces. The effective aperture of a parallel-sided beam is reduced by refraction from a to b in Fig. 9, and the magnification, defined as a/b, is

$$M = \cos P \sec(P + D) \quad . \quad . \quad . \quad . \quad (3)$$

which for a constant value of n increases with P. Since the refracted beam has a smaller effective aperture than the transmitted beam, the width of the response curve at the receiver position is greater than in the direct transmission tests by the factor M, and the error in locating the beam direction increases by the same factor. The estimated error in n as given in eqn. (2) must therefore be modified by including the dependence of δD on P, namely

$$\delta D = M(\delta D)_0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where $(\delta D)_0$ is the error in determining beam directions over a direct transmission path. The combination of eqns. (2)–(4) gives

$$\delta n = \cot P(\delta D)_0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

which shows the rather interesting result that the actual error in the refractive index does not depend on the value of the index itself.

A second factor which must be considered is the reflection at each of the prism surfaces, and the voltage-reflection coefficient at the incident face and at the hypotenuse may be calculated from the Fresnel equations. From these formulae, the percentage of the incident power transmitted through the prism is found to fall as the prism angle increases, becoming zero when internal reflection occurs at the hypotenuse for all prism angles exceeding 38·7°. At this critical angle the refracted beam emerges in a direction parallel with the hypotenuse.

The choice of a suitable prism angle is governed almost entirely by the need to ensure satisfactory transmission of the incident power through the prism, and 30° is about the largest angle that can be usefully employed. The estimated error in the refractive index for $(\delta D)_0 = 0{\cdot}25°$ is 0·008, i.e. about 0·5%, which compares reasonably well with the errors arising in other methods of measurement. Accordingly, three right-angled prisms with a nominal prism angle of 30° were machined from polystyrene; their hypotenuses measure 45·5, 31·1 and 21·5 cm. The thickness of the prisms was made about 5% greater than the plate spacing in the hope that the weight of the top plate pressing on them would ensure a good contact between the dielectric and the plates. As will be seen in Section 7.2, this hope has not been realized. Cavity measurements on a sample cut from the same sheet as the prisms gave the value 1·59 for the refractive index of the polystyrene.

(7.2) Initial Measurements with the Prisms

The first measurements with the prisms were carried out using the horn transmitting aerial and the waveguide receiving aerial. The refractive index calculated from the centre of gravity of the response curve for the largest prism is 1·53, about 4% lower than the value obtained by the cavity measurements; the error is eight times that estimated in Section 7.1. The response curve is neither smooth nor symmetrical and shows ripples of a kind similar to those observed during the tests on the reflecting barrier. The responses obtained with the other prisms differ in detail but have the same general characteristics and lead to the same value of refractive index. Further experimental work was therefore carried out to determine the causes of the error in the refracted beam direction and of the beam distortion. Since these effects seem to be largely independent of the prism size, the additional measurements were restricted to the largest prism.

The prism was first inserted in each of the possible positions equivalent to the arrangement of Fig. 9, and this gave results consistent with that already obtained. Changes in the aerial penetrations did not exert any appreciable effect on the deviation, nor did the arrangement of the absorbing spacers. Measurements were then made for a range of angles of incidence of up to 20° from the normal to the input face of the prism, and gave deviations consistent with the refractive index value of 1·53.

The results of these tests suggested that the source of the error lay in the prism and not in the actual instrument. Possible causes considered were higher-order modes of propagation in the dielectric, contact effects between the dielectric and the plates, and diffraction at the edges of the prism. Experiments of various kinds showed that the first and last of these did not contribute errors of the magnitude being sought.

Theoretical considerations of the kind discussed in relation to dielectrics in rectangular waveguides[13] suggested that errors in the refractive index resulting from small gaps should be negligible. Failure to find any other explanation for the error in the measured refractive index led to a decision to confirm this by direct experiment. The top plate was progressively raised by shims of thickness 0·004 in placed at the prism edges, and the deviation was measured for each gap width. The effective refractive index, calculated from the deviations, is plotted against the gap thickness in Fig. 10, the point B corresponding to absence of shims. The refractive index was also measured when the top plate was weighted, again in the absence of shims; the value obtained was plotted on the curve, extrapolated as in the Figure, to give point A. From this it appears that weighting the top plate closes the gap by 0·005 in, but in fact no physical gap of this magnitude

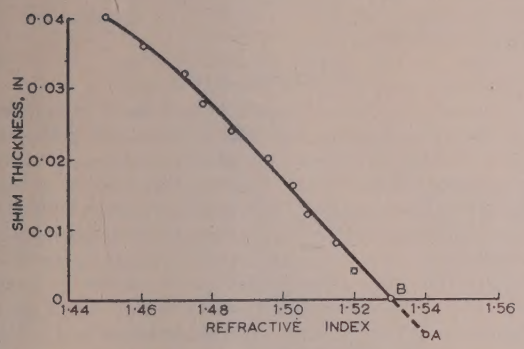


Fig. 10.—Dependence of effective refractive index of prism on thickness of air-gap between dielectric and plates.

The gap was produced by inserting shims of the thickness shown at the prism edges.

exists even when the top plate is unweighted. The correct value of the refractive index of polystyrene appears at a gap reduction of 0·03–0·04 in—far in excess of any gaps occurring. The slope of the curve in Fig. 10 is in reasonable agreement with the theory already referred to.

Experience with parallel-plate transmission lines has shown that for certain artificial dielectrics improved consistency of results was obtained by covering the specimen surfaces with tin foil.[7] This was now tried for the prism, adhesion of the foil being secured by a film of petroleum jelly, and the measured value of the deviation was found to correspond to the correct refractive index, namely 1·59. Check measurements on all three prisms in a wide variety of positions confirmed that a tin-foil coating made the deviations agree consistently with the values predicted for a refractive index of 1·59. Small gaps between the foil and the spectrometer plates allowed a weak signal to leak through to the position directly opposite to the transmitter horn but caused no alteration in the main beam direction.

The use of tin foil to coat the prism surfaces thus permits accurate measurements of the refractive index to be made. No explanation has yet been found for the departures of the deviations from the expected values when the tin foil is absent.

(7.3) Distortion of the Refracted Beam

The ripples on the refracted-beam response curves are similar in form to those for the responses in the reflected-beam test. In that test, the ripples arose from interaction between the reflected beam and a transmitter side-lobe, and it is therefore probable that a similar interaction applies in the prism measurements. The geometry of the arrangement used excludes the possibility that the interfering signal arises from a side-lobe, but interference may occur as a result of diffraction at the prism corners or reflection at the side BC of the prism (Fig. 9). A theoretical analysis of the ripple period to be expected as the result of an interfering signal showed that it was possible to predict the direction from which this signal was arriving, ant this considerably simplifies the problem of determining its origin.

In all the tests of practical interest the wanted signal arrives from the centre of the spectrometer, i.e. in the direction OR, in Fig. 11. Over a small range, the phase fronts of this signal are planes normal to OR and may be shown as the lines XX and X′X′. If the distance R_1R_3 equals the free-space wavelength, λ, the phase of the wanted signal is the same at all points on both XX and X′X′. Suppose that the interfering signal arrives from a direction making an angle ϕ with OR_1 and that YY and Y′Y′ are the phase fronts for which this signal has the same phase as does the wanted signal for the fronts XX and X′X′. At each of

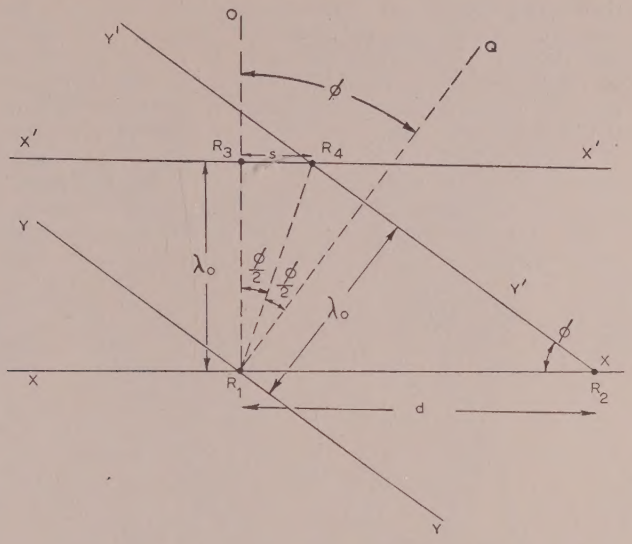


Fig. 11.—Diagram for analysis of interference effects.

the points R_1, R_2 and R_4 the two signals are in phase, and ripple maxima will therefore occur at these points. Since OR_1 is a radius of the spectrometer, the receiving aerial moves in a direction perpendicular to it, provided that the penetration remains constant, and so for small distances may be considered to move on a line parallel to XX. The distance d, which must be traversed between ripple maxima, is obviously equal to R_1R_3, and from the Figure

$$d = \lambda \operatorname{cosec} \phi \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

The receiver movement is recorded in degrees, but since 1° at the circumference corresponds to 1 cm, the distance d can be obtained directly from the response curve and then ϕ may be calculated from the above equation. However, this does not yield any information as to the side of OR_1 on which the interfering signal lies. If now the receiver is moved inwards one wavelength by means of the penetration adjustment, the ripple pattern is shifted laterally by a distance s, as shown, for example, by the displacement of the maximum at R_1 to R_4. From the geometry of the Figure

$$s = \lambda \tan (\phi/2) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

The shift is towards the side of OR_1 from which the unwanted signal arrives.

The source of the distortion in the refracted-beam responses may be discovered with the help of the above theory. Portions of two responses taken with the large prism are plotted in Fig. 12;

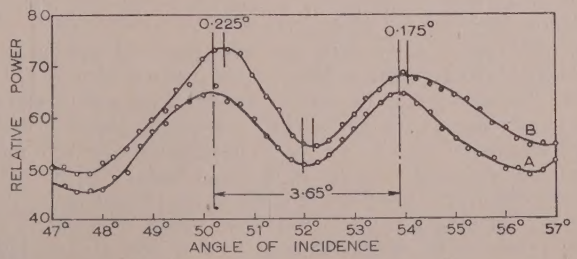


Fig. 12.—Enlarged portions of response curves observed for the beam refracted by the large prism.

Curve B corresponds to a receiving aerial penetration on wavelength greater than that for A.

they correspond to two receiver penetrations differing by one wavelength. The ripple period is 3·65°, giving $d = 3{\cdot}65$ cm so that $\phi = \text{arc sin } (1{\cdot}25/3{\cdot}65) = 20{\cdot}0°$. The angular shifts of two maxima and one minimum when the waveguide receiver penetration is increased by one wavelength are 0·255, 0·200 and 0·175° respectively, giving from eqn. (7) values of ϕ equal to 20·5, 18·5 and 16·5° respectively. In view of the small angular shifts involved, the agreement is remarkably good and demonstrates the accuracy obtainable in measuring changes in direction.

The response curves of Fig. 12 were taken with the prism positioned as in Fig. 13; the direction of the shift in the ripple

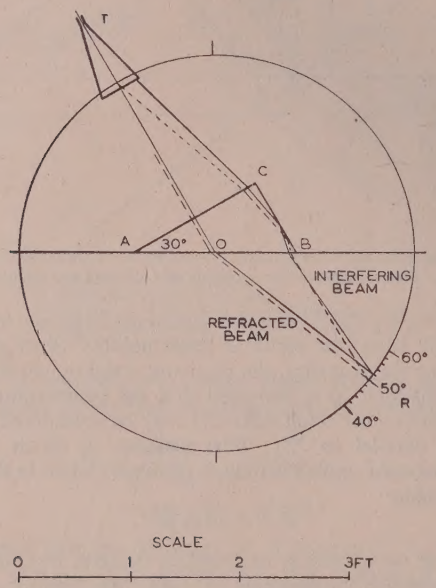


Fig. 13.—Scale drawing of prism location to determine origin of interfering signal.

pattern on increasing the receiver penetration shows that the interfering beam is arriving from the right of the main beam, and an accurate drawing of these directions has been made. This makes it quite evident that the origin of the interference is the total internal reflection occurring at the prism side BC. Detailed analyses of other responses obtained under similar conditions show that this total internal reflection is the principal cause of the distortion. Experimental confirmation of this was obtained by observing that, when absorbing wedges are positioned so as to eliminate the transmitter radiation illuminating the side BC, the ripples on the responses disppear.

(8) TESTS WITH THE HORN AS RECEIVING AERIAL

Most of the measurements already described were repeated with the horn as receiving aerial in place of the waveguide. The principal benefits arising from this substitution are the improved accuracy in locating the direction of the received beam (resulting from the reduction in the receiving-aerial beam width) and the partial elimination of interference phenomena (because of the discrimination of the horn against radiation arriving from directions well removed from that of the main beam). In particular, measurements of the refractive index of the prisms (the surfaces again been coated with tin foil) gave values of 1·583, 1·587 and 1·589 for the prisms with 45·5, 31·1 and 21·5 cm hypotenuses respectively. The error in the refractive index is within the limits predicted in Section 7.1, and shows that beam directions can in fact be determined to within the 0·25° error which was assumed. The ripples on the response curves for the refracted beams are completely eliminated by using the horn as receiver, since its angular response is sufficient to suppress the signal reflected from the prism side.

(9) CONCLUSIONS

The parallel-plate spectrometer as designed and constructed enables the directions of reflected or refracted beams to be measured with errors of less than 0·25°. Changes in direction can be measured to within 0·1° or less. Fluctuations in received signal level may be caused by interference between an unwanted signal such as a transmitter side-lobe and the main beam, but may be avoided by using a sufficiently directive receiving aerial.

The refractive index of a dielectric prism may be measured to within 0·5% provided that the top and bottom surfaces of the prism are coated with tin foil. Failure to use a tin-foil coating leads to an error of about 3%, no explanation being apparent for this discrepancy.

(10) ACKNOWLEDGMENTS

The authors are indebted to Dr. A. W. Adey for providing theoretical results on the diffraction of dielectric cylinders; to Mr. J. S. Seeley for assistance in the preparation of the diagrams; to Mr. C. C. Mills, Superintendent of the Electrical Engineering Workshop, for many helpful suggestions in the design of the instrument; and to Mr. R. Moore for the care and accuracy with which the instrument was made.

The work described was carried out as part of the programme of the Radio Research Board. This paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research.

(11) REFERENCES

(1) CULSHAW, W.: 'The Michelson Interferometer at Millimetre Wavelengths', *Proceedings of the Physical Society*, 1950, **63B**, p. 939.

(2) CULSHAW, W.: 'The Fabry–Perot Interferometer at Millimetre Wavelengths', *ibid.*, 1953, **66B**, p. 597.

(3) CULSHAW, W.: 'A Spectrometer for Millimetre Wavelengths', *Proceedings I.E.E.*, Paper No. 1445 M, March, 1953 (**100**, Part IIA, p. 5).

(4) BROWN, J., and JACKSON, W.: 'The Properties of Artificial Dielectrics at Centimetre Wavelengths', *ibid.*, Paper No. 1699 R, January, 1955 (**102** B, p. 11).

(5) COCHRANE, C. A.: 'An Experimental Verification of the Theory of Parallel-Plate Media', *ibid.*, Paper No. 896 R, March, 1950 (**97**, Part III, p. 72).

(6) BRADY, J. J., PEARSON, M. D., and PEOPLES, S. R.: 'Reflection of Microwaves from Metal Plate Media', *Journal of Applied Physics*, 1952, **23**, p. 964.

(7) EL-KHARADLY, M. M. Z.: 'Some Experiments on Artificial Dielectrics at Centimetric Wavelengths', *Proceedings I.E.E.*, Paper No. 1700 R, January, 1955 (**102** B, p. 17).

(8) RUZE, J., and YOUNG, M.: 'Experimental Determination of the Reflection Coefficient of Metal Plate Media', *Journal of Applied Physics*, 1951, **22**, p. 277.

(9) BROWN, J.: 'Microwave Lenses' (Methuen, 1953).

(10) MONTGOMERY, C. G.: 'Technique of Microwave Measurements' (McGraw-Hill Book Co., 1947).

(11) TULLER, W. G., GALLOWAY, W. C., and ZAFFARANO, F. P.: 'Recent Developments in Frequency Stabilization of Microwave Oscillators', *Proceedings of the Institute of Radio Engineers*, 1948, **36**, p. 794.

(12) STRATTON, J. A.: 'Electromagnetic Theory' (McGraw-Hill Book Co., 1941).

(13) MONTGOMERY, C. G., DICKE, R. H., and PURCELL, E. M.: 'Principles of Microwave Circuits' (McGraw-Hill Book Co., 1948).

High quality, stable characteristics, robust construction, proven performance, are inherent features of the quartz crystal units developed and manufactured by S.T.C.

Crystals in hermetically sealed cans to meet the rigid Inter-Service specification; crystals in evacuated glass envelopes with wire ends or with valve base; G.T. Cut crystals for absolute frequency standards; high frequency Overtone crystals for V.H.F. applications . . . these are the principal types in the S.T.C. range.

There is almost certain to be an S.T.C. quartz crystal unit to meet your specification . . . if not, our engineers are ready to discuss your special requirements.

The quartz crystal units illustrated are in hermetically sealed cans equivalent to the Inter-Service Style "D" or the American HC6U. This unit has been widely adopted for use in modern equipments.

Standard Telephones and Cables Limited

Registered Office: Connaught House, Aldwych, W.C.2

TELEPHONE LINE DIVISION · NORTH WOOLWICH · LONDON · E.16

PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

Part B. RADIO AND ELECTRONIC ENGINEERING (INCLUDING COMMUNICATION ENGINEERING) MAY 1956

CONTENTS

Bibliographical References.—It is requested that bibliographical reference to an Institution paper should always include the serial number of the paper and the month and year of publication, which will be found at the top right-hand corner of the first page of the paper. This information should precede the reference to the Volume and Part.

Example.—Smith, J.: 'Reflections from the Ionosphere,' *Proceedings I.E.E.*, Paper No. 3001 R, December, 1954 (102 B, p. 1234).

Published by The Institution, Savoy Place, London, W.C.2. *Telephone:* Temple Bar 7676. *Telegrams:* 'Voltampere, Phone, London.'
Printed by Unwin Brothers Limited, Woking and London.